Manuscript Details

Manuscript number	YPTSP_2019_561_R1
Title	Is markerless, smart phone recorded two-dimensional video a clinically useful measure of relevant lower limb kinematics in runners with patellofemoral pain? A validity and reliability study.
Article type	Research Paper

Abstract

Objectives: Investigate the validity and reliability of markerless, smart phone collected, two-dimensional (2D) video, analysed using the 'Hudl technique' application, compared to three-dimensional (3D) kinematics during running, in participants with patellofemoral pain (PFP). Design: Validity/reliability study Setting: Biomechanics laboratory Participants: Males/females with PFP (n=21, 10 males, 11 females, age 32.1 months [±12.9]). Main Outcome Measures: Manually synchronised 2D and 3D measurement of peak hip adduction (HADD) and peak knee flexion (KFLEX) during running. Results: 2D and 3D measures of peak KFLEX (p=0.02, d=1.13), but not peak HADD (p=0.25, d=-0.27), differed significantly. Poor validity was identified for 2D measurement of peak HADD (ICC 0.06, 95% CI -0.35, 0.47) and peak KFLEX (ICC 0.42, 95% CI (-0.10, 0.75). Moderate intra-rater reliability was identified for both variables (ICC 0.61-65), alongside moderate inter-rater reliability for peak KFLEX (ICC 0.71) and poor inter-rater reliability for peak HADD (ICC 0.31). Conclusions: Measurement of peak HADD and KFLEX in runners with PFP using markerless, smart phone collected 2D video, analysed using the Hudl Technique Application is invalid, with poor to moderate reliability. Investigation of alternate 2D video approaches to increase precision is warranted. At present, 2D video analysis of running using Hudl Technique cannot be advocated.

Keywords	Patellofemoral Pain; Running; Kinematics; Validity
Taxonomy	Musculoskeletal System, Biomechanics of Gait

Corresponding Author	Bradley Neal
Corresponding Author's Institution	Queen Mary University of London
Order of Authors	Bradley Neal, Simon Lack, Christian Barton, Aleksandra Birn-Jeffery, Stuart Miller, Dylan Morrissey
Suggested reviewers	Chris Bramah, Bart Dingenen

Submission Files Included in this PDF

File Name [File Type]

Cover Letter_PTiS_2D3D_Revision.docx [Cover Letter]

Reviewer Response Document.docx [Response to Reviewers (without Author Details)]

Highlights.docx [Highlights]

Title Page_Revision.docx [Title Page (with Author Details)]

2D 3D Manuscript_Revision_FINAL.docx [Manuscript (without Author Details)]

Conflict of Interest.docx [Conflict of Interest]

Ethical Approval.docx [Ethical Statement]

Funding.docx [Author Agreement]

2D 3D Manuscript_Revision_Tracked.docx [Supporting File]

Research Data Related to this Submission

There are no linked research data sets for this submission. The following reason is given:
The data that has been used is confidential

Barts and The London
School of Medicine and Dentistry

Sports and Exercise Medicine
Bart's and the London School of Medicine and Dentistry
Queen Mary, University of London

Thursday, 06th February 2020

Dr. Lee Herrington, PhD

Editor, Physical Therapy in Sport

Senior Lecturer in Sports Rehabilitation, School of Health Sciences, University of Salford, Salford, UK.

Re: Is two-dimensional video a clinically useful measure of relevant lower limb kinematics in runners with patellofemoral pain? A validity study.

Dear Dr Herrington,

Thank you for considering re-reviewing our paper, which we felt was appropriate for submission to PTiS given your journals' publication of previous works in this field (Dingenen et al 2017, 2018, 2019). We are confident that we have appropriately addressed or offered an appropriate rebuttal to all reviewer comments and feel that our manuscript is certainly stronger after the kind and constructive comments of all reviewers.

This manuscript represents the result of many months of work investigating the validity and reliability of the analysis of running kinematics using 2D video, compared to 3D motion capture, in a group of runners with patellofemoral pain. Our paper identifies that the commercially available HUDL application, which is free at point of access and used widely amongst clinicians, is invalid and does not accurately predict either 3D peak hip adduction or peak knee flexion. We have attempted to discuss why this negative finding is in conflict with previous works and have made appropriate suggestions for future works to improve upon this. We have made greater light of our reliability data and added a clinical implications section on the advice of the reviewers.

All of the authors have read and concur with the final content in the manuscript. The material within has not been and will not be submitted for publication elsewhere except as an abstract. Neither myself nor any of the other authors have any competing interests. All authors made substantial contributions to the conception, design and delivery of the review and all authors contributed to the final manuscript preparation, before I gave final approval for this version to be submitted.

With best wishes,

Dr Bradley Stephen Neal, PhD
Research Fellow
Centre for Sports and Exercise Medicine
William Harvey Research Institute
Barts and the London School of Medicine and Dentistry
Queen Mary, University of London

e: b.s.neal@qmul.ac.uk
m: 07732430064
t: @Brad_Neal_07

A reviewer response document is presented, with a point by point acceptance/rebuttal to the reviewers' constructive comments. Elements of the manuscript have been copied here for ease of reviewing, with changes reflected by italicised and underlined text.

Reviewer 1

Dear authors,

Thank you for submitting this interesting manuscript. The content of the paper is clinically useful, and relevant for the readership of Physical Therapy in Sport.

Response: We thank the reviewer for their time in providing us with such constructive comments. We are glad that you feel our paper is clinically relevant and appropriate for the journal.

The main remark I have is that this paper shows actually that the method is not reliable, while the authors mainly focus on the outcomes of the validity. While the validity outcomes are very interesting for any clinician, the first premise of a measurement is that the measurement is reliable, and this is not the case, for various reasons. If your measurement is not reliable, then it makes no sense to focus in first place on the validity. So in general, I would advise giving more attention to the lack of reliability, prior to focusing on the validity analysis.

Response: We thank the reviewer for this interesting comment. We would agree that we should make more of our reliability data, which demonstrates poor reliability as well as poor validity. However, the over-arching aim of the study was to determine if a 2D video method can accurately predict 3D kinematics (validity). As such, we have written the manuscript such that the validity outcomes (accuracy) come ahead of the reliability outcomes (repeatability).

We have amended the title of our manuscript to reflect both the method of our 2D video recording and analysis, and the reliability element of the study. It now reads as follows:

"Is *markerless*, *smart phone recorded* two-dimensional video a clinically useful measure of relevant lower limb kinematics in runners with patellofemoral pain? A validity *and reliability* study"

Another general comment is that across the whole manuscript, a lot of sentences don't start with a capital letter. Please correct this, even though this is only a remark on writing style.

Response: We apologise for this error and confirm it has been corrected throughout the manuscript.

Please also include line numbers in the revised version; this makes it easier for us as a reviewer to provide feedback.

Response: We are not sure how this error has occurred, as the journal submission system applies line numbers to the eventual .pdf and the file we have from the system includes line numbers.

In the following lines, I provide my feedback on a point by point basis. The feedback being formulated should be interpreted as positive feedback to improve the final quality of the manuscript.

Abstract

Objective: It's better to included immediately the specific type of 2D analysis in your objective (Hudl technique). Across the whole manuscript, I believe it's important to mention that the methodology being used in this study is not reliable and valid, so no overgeneralisation should be made to any 2D measurement.

Response: This is a very fair comment. Throughout the manuscript we have amended this description so that the reader is aware that we used a specific 2D analysis method, so as to avoid over-generalisation to all forms of 2D assessment. The objective in the abstract now reads as follows:

"Objectives: Investigate the concurrent validity and reliability of *mobile phone collected,* high frame rate two-dimensional (2D) video, *analysed using the 'Hudl technique' application,* compared to three-dimensional (3D) kinematic motion capture during running, in participants with patellofemoral pain (PFP).

Design: you also measure reliability (as I mentioned already, I think this is highly important).

Response: We agree and have included 'reliability' in both the title and the study design section of the abstract.

Participants: There is a large range in age, height and weight. Please be more specific in your description of the recruitment process and inclusion and exclusion criteria in the method section.

Response: We endeavoured to recruit participants reflective of a heterogenous patellofemoral pain cohort and no exclusion criteria were applied to either height or body mass. We did have an ag exclusion and apologise for omitting this, but have now added this information to the manuscript. This now reads as follows:

"Participants *below the age of 18 or over the age of 50*, *or those* with *traumatic symptoms*, patellofemoral instability, tibiofemoral pathology or other concomitant pathology were excluded."

I have difficulties to understand based on the description in the paper how the 2D and 3D analysis could have been synchronized. As far as I can understand, the phone was not connected with the 3D system, so it's impossible to synchronize data collection. Please explain.

Response: Our 2D and 3D data collection was manually synchronised as described in the methods section. We have added the 'manual' descriptor to the abstract for greater clarity and made it clear that the 2D data were analysed independent of the 3D data. This now reads as follows:

"Videos from successful trials were subsequently imported into the Hudl Technique application (Hudl, Agile Sports Technologies Inc., Nebraska, USA) *and analysed independently of the 3D data*."

Introduction, second paragraph: please include also other intervention studies showing an effect on hip adduction and symptom improvement, next to the step rate retraining intervention study.

Response: We have cited both the work of Noehren (2011) & Willy (2012), which specifically report changes to hip adduction in relation to symptom improvement as requested.

Why would it be important to you that validity would be different in a clinical population? Was the variability indeed different between your study and the other studies?

Response: We agree that our argument for investigating a clinical population in the introduction could have been stronger. We have amended this, which now reads as follows:

"Whilst both of these studies reported their methods to be reliable (ICC 0.90-0.99), *given that runners with PFP demonstrate differing kinematics compared to matched controls, (B. S. Neal, et al., 2016)* the investigation of asymptomatic runners limits the external applicability to clinical populations."

In the introduction, only hip adduction is discussed. Knee flexion is not mentioned, while this is one of the two "relevant" outcome parameters, according to you. While there can be a theoretical rationale to included knee flexion, I am not sure there is any strong retrospective and prospective evidence showing that patients with patellofemoral pain run with less or more knee flexion (Ceyssens et al. 2019, Neal et al. 2016). A lot of studies fail to find differences in knee flexion. Please reflect on this.

Response: We would agree with the reviewer that the link between hip adduction and PFP is certainly stronger than the link with knee flexion. Whilst we had made links between peak knee flexion and PFP to warrant our investigation of the valid measurement of this variable, we have strengthened this narrative, which reads as follows:

"Peak knee flexion is also a variable of interest in runners with PFP. *It* is reported to *correlate with* patellofemoral joint stress (Lenhart, Thelen, Wille, Chumanov, & Heiderscheit, 2014) *and is also associated with kinesiophobia, with females with PFP demonstrating lower peak knee flexion angles during stair descent. (de Oliveira Silva, Barton, Pazzinatto, Briani, & de Azevedo, 2016) Altering* peak knee flexion may be associated with symptomatic improvements after a step rate retraining intervention. (Bradley S Neal, Barton, Birn-Jeffrey, Daley, & Morrissey, 2018)"

Method: Please be more specific on the running characteristics of the participants. To be eligible, participants should have pain during at least one of the activities mentioned in the method section. So, theoretically, it could be possible that participants did not have knee pain during running? What is the definition of a runner in your cohort? How far did they run? What speed? Did they stop running? No information is provided on this topic. Please include more detailed information on the participants (runners?) in this study.

Response: We thank the reviewer for this point. We did not recruit a running specific sample for this study, but instead focussed on recruiting a heterogeneous sample of participants with PFP who experienced pain during multiple tasks, reflective of wider clinical practice. We felt as though this was more in line with the overall aim of the study, which was to determine accuracy of measurement, as opposed to making any inferences regarding the association between biomechanics & symptoms, where we agree a running specific sample would be essential. This is why we chose to collect a Tegner score, as a reflection of overall participant activity level.

A major limitation of the study, that is currently not addressed, is that 2D kinematic data were assessed with an iPhone. The problem when using a Hudl Technique app on an iPhone, compared to for example a tablet or a PC, is that the screen is a lot smaller. The smaller the screen, the more difficult it will be to make a correct placement of the anatomical points to define the angles. The previous studies drew their angles on a computer screen.

Response: We thank the reviewer for this important point and apologise for a minor oversight on our part. Whist the 2D videos were recorded using an iPhone 6, 2D data analysis was conducted with a 5th generation iPad, with a 10.2" screen. We have included this information in the manuscript, which now reads as follows:

"Videos from successful trials were subsequently imported into the Hudl Technique application (Hudl, Agile Sports Technologies Inc., Nebraska, USA) for analysis. *2D data analysis was completed using a tablet device with a 25.9cm screen (5th generation iPad, Apple Corporation, California, USA).* Two independent 2D angles, hip adduction (HADD) and knee flexion (KFLEX) were identified.

We also fully agree with the reviewer with respect to screen size and have made greater comment on this in the discussion section.

The frontal videos were made from a distance of 6 meters. Based on the figure, it seems that a zoom function was used, is this correct? Please reflect on this.

Response: The reviewer is correct, when analysing the 2D videos, use of the zoom function was permitted at the discretion of the analyser, to the point where bony landmarks were required for 2D angle extraction were best visualised. We have included this key piece of information in the manuscript. This now reads as follows:

"*Use of the zoom function within the Hudl technique application was permitted at the discretion of the analyser, to ensure optimal visualization of the relevant anatomical landmarks*".

Please explain how the 2D and 3D system could have been synchronized. I don't believe this this was indeed the case based on the information in the manuscript.

Response: The 2D and 3D systems were manually synchronised using a verbal countdown, initiated by a member of the research team. As the 2D and 3D data were analysed independently and we did not use a time point from one data set to identify an equivalent time point in the other, we do not see how this could have impacted our results.

2D kinematic analysis: it is argued that hip adduction and knee flexion are independent angles, but what is the evidence that this is true?

Response: We would agree with the reviewer that these kinematic angles are indeed coupled and thus not truly independent of one another. However, in order to investigate the agreement between 3D and 2D measurement, it was necessary for us to analyse these angles as independent kinematic variables within a single plane.

Please clarify how the exact angles were calculated. For example, the peak knee flexion angle shown in figure 2 is not the same angle as shown in the table. Please clarify how the angles from 2D were derived to the specific angles being used in the data analysis. The same for the other angles.

Response: We thank the reviewer for raising this important point. We have added this key information to the manuscript for both variables, which now reads as follows:

"HADD was determined using methods described by Dingenen et al, where the contralateral pelvic drop (CLPD) angle is added from the femoral adduction (FADD) angle. (Bart Dingenen, et al., 2017) CLPD angle was defined as the angle formed by a horizontal line from the stance limb anterior superior iliac spine (ASIS) (referenced from the laboratory floor) and the swing limb ASIS (see figure 2). FADD angle was defined as the angle formed by a horizontal line from the stance limb ASIS (referenced from the laboratory floor) and the centre of the stance limb tibiofemoral joint (an estimation of the knee joint centre) (see figure 2). *Within the Hudl technique application, the tool reflects an angle relative to 90˚ and the FADD angle was therefore determined by subtracting the angle produced by the tool from 90˚*. Infrared ASIS and PSIS markers used for 3D kinematic data collection were visualised to determine the location of these anatomical landmarks on 2D video.

KFLEX was defined as the angle formed by a line drawn from the stance limb greater trochanter to the lateral femoral condyle and a second line drawn from the stance limb lateral femoral condyle to the stance limb lateral malleolus (see figure 2). *Within the Hudl technique application, a vertical line in the sagittal plane is reflective of 180˚ and the KFLEX angle was therefore determined by subtracting the angle produced by the tool from 180˚*. For both variables, a peak angle was estimated, determined to be when the participant reached the peak of mid-stance, manually defined as the point where maximal foot contact had occurred and no upward/downward motion was occurring. (Maykut, et al., 2015)"

For the validity analysis, the mean of 5 trials was used. However, for the reliability analysis, only one trial was used to calculate reliability? Is this correct? Normally, the same method should be used to calculate the reliability and the validity. Taking one or 5 steps can make a difference in this kind of measurements.

Response: Thank you for your comment and you are correct, a different methodology was used to investigate validity compared to reliability. This decision was made apriori and for the validity measure, it was agreed that a mean of 5 trials for both 3D and 2D would be compared to determine agreement between systems. However, as reliability is a question of the repeatability of drawing only a 2D angle, with comparison between raters rather than systems, we did not feel the need to compare mean pooled data. We have however added a reflection on this point in the discussion, which reads as follows:

"*Finally, a single video, rather than mean pooled data, were used for the investigation of reliability, differing from the investigation of validity. Whilst this decision was made apriori, analysis of mean pooled data may have yielded different reliability results.*"

Table 4: Please add the size of the screens, as previously mentioned.

Response: This has been added as requested.

For the 2D method, no reflective markers were used to visualize the anatomical landmarks. However, in the method section, it is mentioned that anatomical markers were placed on the lateral femoral condyles (and other places). This is not visible on the figures? Please explain why no markers were visible for the 2D analysis while these markers were used for 3D analysis at the same time.

Response: the markers used for the 3D motion capture are infrared (rather than retroreflective) and are also small (10mm). This style of marker was also placed only on the ASIS/PSIS/lateral calcaneal process and 5th metatarsal head. Rigid clusters of 4 markers were also placed on the thigh and shank respectively. We have added this information to the manuscript for clarity, which reads as follows:

"Kinematic data were collected during running using a four-camera, infrared motion analysis system using Odin software (CX-1, Codamotion, Charnwood Dynamics Limited, Leicestershire, UK), *sampling at 200Hz.* (Lack, et al., 2014) 24 infrared markers; eight individual markers (10mm) and four rigid clusters of four markers (140mm), were placed *adhering to* the CAST protocol. (Cappello, Cappozzo, La Palombara, Lucchetti, & Leardini, 1997) *Individual markers were placed on the anterior superior iliac spine (ASIS), posterior superior iliac spine (PSIS), lateral calcaneal process and head of 5th metatarsal, with rigid clusters placed on the mid-point of each thigh and shank segment. Foot markers were placed on the participants shoe as an estimation of the anatomical location, given the potential for barefoot running to effect running kinematics. (Hall, Barton, Jones, & Morrissey, 2013)*"

The femoral condyle and ankle malleoli markers are virtual and as a result are not visible in the 2D videos.

Last paragraph discussion: the argument for measuring HIR is made. Personally, I would not be that confident measuring HIR with the marker set being used, and especially not with the IMU's, given their measurement errors.

Response: We agree with the reviewer in that even 3D measurement of transverse plane hip data has its limitations. Upon reflection, we also feel that bringing IMUs into the discussion at this point distracts from the main theme of the discussion, which is the lack of validity and reliability in our methodology. As we have now added a 'clinical implications' paragraph at the request of reviewer 2, which we feels strengthens the end of our discussion nicely, we have decided to delete the paragraph in question.

Conclusion: A lot more attention should be given to the fact that your measurement is not reliable.

Response: we agree with the reviewer and have restructured our conclusion accordingly. This now reads as follows:

"*Measurement of both peak HADD and KFLEX in runners with PFP using mobile phone collected, high frame rate 2D video, analysed using the Hudl Technique Application is invalid, with poor to moderate reliability.* This may be attributed to the employed 2D video or statistical methodologies, but could also be explained by the increased variability in running kinematics of runners with PFP. Further investigation of methodologies with increased precision is warranted, aiming to improve the ability of high frame rate 2D video to accurately predict 3D kinematics in the clinical setting. At present, clinical gait analysis conducted using the Hudl Technique application should be interpreted with caution, as the accuracy *or reliability* of 2D measurement cannot be guaranteed.

Reviewer 2

The study aims to investigate the validity and reliability of a commonly used 2D video analysis method (Hudl technique via an IPhone) compared to 3D measurements of peak hip adduction and knee flexion angle. The methodology appears robust and appropriate for the study question. The results from the study highlight poor validity of 2D measurements when compared to 3D, as well as only moderate intra and interrater reliability. Authors discuss how the lack of construct validity could be explained by the poor precision of digitising kinematics using touch screen technology compared to computer based 2D packages.

I feel this study has some important messages for clinical practise, that being that 2D assessment measures using common clinical methods is not a valid measurement tool when investigating 3D running kinematics.

However, I feel this message could be more clearly made within the discussion section of the paper for a stronger manuscript. Otherwise this is a very well written manuscript and interesting study.

Response: We thank the reviewer for their time in providing such constructive comments.

Specific Comments

Introduction

Paragraph 1: Maybe include some more figures here to highlight the significance of PFP?

Response: some injury prevalence figures have been added as requested. This now reads as follows:

"Recreational running is a common form of exercise (Linton & Valentin, 2018) associated with both positive health benefits (Lee, et al., 2017) and high rates of musculoskeletal injury *(19-94%)*. (Saragiotto, et al., 2014) The knee is reported to be the most prevalent joint involved in running-related musculoskeletal injury, (Linton & Valentin, 2018; Taunton, et al., 2002) with patellofemoral pain (PFP) the most prevalent diagnosis *(17%)*. (Taunton, et al., 2002)"

Paragraph 4: I would like to see a clearer explanation as to why hip internal rotation may impact 2D measurements of 3D kinematics. You lead to this in the discussion with the sentence "Ortiz et al (219) hypothesised that transverse plane hip motion may affect the accuracy of 2D measured frontal plane hip kinematics." This a key point that justifies some of your statistics which I think need to me made clearer for the reader.

Response: Thank you for highlighting this oversight. We have added an amended version of this sentence to the start of this paragraph, to add clarity to your rationale to investigate the potential for hip internal rotation to act as a confounding factor. This now reads as follows:

"Previous studies investigating construct validity for 2D video to measure peak HADD have *not identified* optimal agreement between 2D and 3D measurement. *Ortiz et al hypothesised that transverse plane hip motion may affect the accuracy of 2D measurement of frontal plane hip kinematics during a jump/land task. (Ortiz, et al., 2016)* Runners with PFP have also been reported to demonstrate increased peak hip internal rotation (HIR) in comparison to controls. (Noehren, Pohl, Sanchez, Cunningham, & Lattermann, 2012; Noehren, Sanchez, Cunningham, & McKeon, 2012; Souza & Powers, 2009a, 2009b; R. W. Willy, Manal, Witvrouw, & Davis, 2012) Transverse plane motion at the hip is coupled with HADD and tibial abduction, referred to in combination as dynamic knee valgus. (Powers, 2010) Determining the impact of this movement direction on the variability observed between 2D and 3D measurement may provide insight into the source of previously reported sub-optimal agreement."

Methods

“Means of the five 2D and 3D trials were calculated and subsequently pooled, leaving one mean 2D and 3D value for each participant for both variables of interest (HADD and KFLEX).”

Comment: sounds confusing, would the following read easier: “The mean of the five 2D and 3D trials were calculated for each participant for both variables of interest (HADD and KFLEX).”

Response: We agree, and this change has been made as requested.

“Finally, a backward linear regression was performed to assess the effect of including 3D peak hip internal rotation (HIR) in a predictive model, with the F change statistic used to determine the significance of 3D HIR as a covariate”

Comment: could you make this a little clearer as to why you did this.

Response: We have attempted to make our statistical rationale for this decision more straightforward. This now reads as follows:

“*Finally, to assess the influence of including 3D peak hip internal rotation (HIR) in a predictive model,* a backward linear regression was performed, with the F change statistic used to determine *if* 3D HIR explains *the hypothesised imperfect agreement between 2D and 3D measurement.*”

Results

Figures 3 & 4: you’ve labelled these as both being HADD, please amend.

Response: We apologise for this error and thank the reviewer for identifying it. This has been amended as requested.

Intra / Intertester reliability: it would ne nice to see the standard error of measurement. My question here is, with moderate reliability, could 2D assessments provide a reliable measurement tool to monitor pre and post intervention effects in kinematics? Including the SEM would highlight whether there is any use in this measure as a measurement tool at all.

Response: this is an excellent suggestion. We have added SEM for both intra- and inter-reliability data as requested. This reads as follows:

Intra-rater reliability

Moderate intra-rater reliability was identified for peak HADD (ICC 0.65 95% CI 0.34, 0.83, *SEM 1.8°*) and peak KFLEX (ICC 0.61 95% CI -0.09, 0.87, *SEM 2.7°*).

Inter-rater reliability

Poor inter-rater reliability was identified for peak HADD (ICC 0.31 95% CI -0.06, 0.64, *SEM 3.1°*). Moderate inter-rater reliability was identified for peak KFLEX (ICC 0.71 95% CI 0.16, 0.89, *SEM 1.4°*).

Discussion

The discussion section focuses on contrasting current findings to previous work as well as limitations and methods for future improvement. I think the results of this study have some strong clinical implications that I would like to see more clearly highlighted. That being the lack of construct validity and only moderate intertester reliability when using this technique and therefore the lack of clinical utility of this method.

Response: we thank the reviewer for this comment and have added a clinical implications sub-section to finish the discussion, aiming to highlight the excellent points made by all reviewers. This reads as follows:

"*Whilst the results of this study suggest that markerless, smart phone collected high frame rate 2D video, analysed using the Hudl technique application, is not a valid or reliable method of determining 3D running kinematics, there are some implications for clinical practice. Rather than being concerned about maximising video frame rate, attention should be given to placing the 2D camera(s) as close to the runner as possible, to increase quality and reduce parallax error potential. This is most easily achieved using a treadmill rather than over ground running. In addition, use of retroreflective markers is encouraged to maximize ease of identifying relevant bony landmarks, especially those that may be obscured by adipose tissue or clothing. Finally, clinicians are encouraged to analyse 2D data using a large screen and with software that allows for increased precision via use of a computer mouse (or equivalent), rather than a smaller tablet with a touch screen, which is likely to yield inaccurate results."*

The point on statistical differences between studies is an interesting and important one. Previous studies (e.g. Maykut) utilised persons correlation coefficient to determine the construct validity of 2D v 3D gait analysis. Looking at your results a similar conclusion could be drawn if just using Persons (r) regarding peak knee flexion. This is an interesting point that I think could be explored/ described in more detail within your discussion to emphasise the difference between the two measurement systems. In particular, how pearsons correlation statistic does not provide an accurate estimation as to the level of agreement between two measurement systems.

Response: We agree with the reviewer that this is an interesting finding. We have added greater detail to this section of the discussion, which now reads as follows:

"A further *potential* explanation for this conflict *is* the statistical methodologies employed. Maykut et al (Maykut, et al., 2015) calculated a Pearson's Correlation Coefficient (r) *which, as a bivariate test,* (George, Batterham, & Sullivan, 2003) may over-estimate the agreement between two variables where data demonstrates a linear trend (McGraw & Wong, 1996). *This is reflected by the high (r) produced by the peak KFLEX data from this study (0.74), versus the low (r) produced by the peak HADD data (0.07)."*

Limitations:

Why did you restrict to just two kinematic parameters? While this is not a problem we do not know if other parameters commonly associated with running related injuries and performance may demonstrate greater reliability. This may also warrant further investigation.

Response: We chose to investigate HADD and KFLEX given their association with running-related PFP. We agree with the reviewer however and have listed this as a limitation, which reads as follows:

"Only two kinematic variables were assessed in this study and it may be that other kinematic variables prove to be both valid and reliable if investigated by future studies."

Reviewer 3

Thank you for the opportunity to review your manuscript investigated the use of 2D video in runners with PFP. This is certainly an interesting and relevant area of study. Please see my comments below for suggestions to improve the manuscript.

Response: We thank the reviewer for their time in providing us with such constructive comments. We are glad that you feel our paper is both interesting and relevant.

Title

I think the title needs to reflect the use of mobile-phone technology rather than simply 2D video and this should be emphasised more clearly throughout the manuscript.

Response: We agree with the reviewer and have changed the title accordingly. It now reads as follows:

"Is *markerless*, *smart phone recorded* two-dimensional video a clinically useful measure of relevant lower limb kinematics in runners with patellofemoral pain? A validity *and reliability* study"

Introduction

I feel the introduction does a good job of providing background and rationale for the study. I would like to see more clarity on the fact that this is a phone-based video analysis in the aims.

Response: We agree with the reviewer's comment and those of reviewer 1. We have endeavoured to make reference to the fact that our 2D video recording and analysis was both markerless and conducted using a smart phone. Our aims now read as follows:

"This study aimed to determine whether clinicians can use a simple, readily available tool to measure important lower limb kinematic variables during running. The primary objective was to investigate the concurrent validity and intra- and inter-rater reliability of *markerless*, high frame rate 2D video, *recorded using a smart phone*, with reference to 3D kinematic motion capture. The null hypothesis was that *smart phone collected* 2D video would not give useful measurements of acceptable accuracy with respect to 3D kinematic analyses and as such, a secondary objective was to investigate the source of any identified disagreement."

Methods

Overall, the methods need an overhaul. The order of information between subheadings is confusing and difficult to find.

Response: Thank you for your comment. We have made amendments to improve the clarity to all aspects of our methodology sections and hope that the reviewer now finds the information more readily accessible.

Line 320: is a 4-camera system adequate for capturing this data? According to Figure 1 two of the cameras would have been virtually redundant in capturing frontal plane data due to their positioning posterior to the participants. Two camera tracking of the marker trajectory in the frontal plane will decrease the accuracy of the data.

Response: The positioning of the 3D cameras is consistent with the methodology used by our group for multiple previous running kinematic analyses, both over-ground and treadmill (Neal et al, 2018 PTiS, Neal et al, 2029, J Biomech). Once laboratory coordinates have been established on the laboratory floor, two cameras (anterior and posterior) to a participant are adequate for collecting gait data given the cone-shaped infra-red emission from the anatomical markers. Therefore, four-cameras are more than adequate for accurate collection of 3D kinematic data using the Odin system.

What was the sampling frequency of the 3D cameras?

Response: Please accept our apologies for omitting this key information. The sampling frequency was 200Hz. We have amended the manuscript as follows:

"Kinematic data were collected during running using a four-camera, infrared motion analysis system using Odin software (CX-1, Codamotion, Charnwood Dynamics Limited, Leicestershire, UK), *sampling at 200Hz*."

What force plate was used and what was the sampling frequency?

Response: We again apologise for failing to state the sampling frequency of the force plate (1000Hz). We had listed the type and model of the force plate used (Type 9281CA, Kistler Corporation, Switzerland), but not at its first mention. This oversight has now been corrected.

Line 354: Is an iPhone 6 still a relevant camera for use in clinical practice? What is the quality of the HS video on this camera compared to those of previous studies (Maykut, Dingenen) and, although a point for the discussion rather than methods, how would this affect the data collected?

Response: This is an interesting point. Both Maykut (60Hz) and Dingenen (50Hz) actually used lower 2D recording frequencies than in our study (240 frames per second). We agree that this is of value to our discussion and have added these data to the comparison table and discussed this in our clinical implications section.

How did you ensure the camera was level? This is important considering that the lab floor was used as a reference for the 2D measures. Was the use of the lab floor necessary?

Response: In-built levelling tools built into the tripods used to mount the 2D cameras ensured that they were level. As the cameras were ensured to be level relative to the laboratory floor, we continue to feel as though this was the easiest way to explain how horizontal lines were drawn within the 2D analysis.

Could HADD be measured using ASIS and the knee rather than the addition of CLPD and HADD?

Response: We agree with the reviewer that it certainly could but endeavoured to ensure that our work was comparable to the previously completed normative studies. This was the methodology used by Dingenen et al, hence our replication.

Why was the frontal plane camera recording at 2.5m and the sagittal plane at 6.5m? What affect might this have had on the error of 2D video measurement?

Response: This is simply a case of laboratory set up. The sagittal plane camera could not be any closer to the force plate to allow the entire participant to be visualised on screen. The frontal plane camera had to be far enough away from the centre of the force plate so that the participant did not run into it during data collection, risking injury to themselves and damage to the equipment.

Figure 1: can you clarify the distance of the starting point? In line 412 you have indicated that participants ran 10m but the starting point of the trial was 5m in line 423

Response: We apologise if this was confusing. Participants ran for a total of 10m with the force plate at the mid-point of this distance. We have made this clearer by stating that participants ran for approximately 10m and that whilst they were instructed to land the affected limb on the force plate, they continued to run through so that no deceleration was occurring during the stance phase within which kinematic data were collected. This now reads as follows:

"The ground-embedded force plate was 5.0 metres from the trial start-point, with participants typically making contact with their fifth step *as they ran through*. Several practice runs were permitted to allow for familiarisation and to ensure adequate force plate contact during a participant's natural running gait *without deceleration*."

3d kinematic analysis: did any data filtering take place?

Response:

2D kinematic analysis: was all analysis undertaken on Hudl on the phone? How big was the screen? How would this affect the accuracy and reliability of marker placement (discussion point)?

Response: Reviewer one also made this point and we apologise for this oversight. Whist the 2D videos were recorded using an iPhone 6, 2D data analysis was conducted with a 5th generation iPad, with a 10.2" screen. We have included this information in the manuscript, which now reads as follows:

"Videos from successful trials were subsequently imported into the Hudl Technique application (Hudl, Agile Sports Technologies Inc., Nebraska, USA) for analysis. *2D data analysis was completed using a tablet device with a 25.9cm screen (5th generation iPad, Apple Corporation, California, USA).* Two independent 2D angles, hip adduction (HADD) and knee flexion (KFLEX) were identified.

We agree that the size of the screen is likely to be a component that affects the subsequent validity of 2D data analysis and have included this in the newly added 'clinical implications' section.

Lines 494 and 500: why were the knee joint and greater trochanter markers estimated from the video rather than using markers? What influence might this have had on reliability and validity?

Response: Our attempt to analyse as clinically applicable a 2D video method as possible meant that we did not use markers for our 2D video analysis. We previously stated that the 3D ASIS markers, visible in the 2D video, were used to aid 2D analysis of CLPD, which we have removed to avoid confusion. This could never have been the case for our 2D analysis of KFLEX as our 3D system has no greater trochanter marker involved and the lateral femoral condyle and lateral malleolus markers are virtual (and therefore not visible).

We fully agree that retroreflective markers will increase the precision (and therefore validity and reliability) of 2D analysis and make strong reference to this in the new 'clinical implications' section, added at the request of reviewer 2.

Line 504: why this point for estimation of peak angle shown to be correct by the 3D data?

Response: We chose this point of estimation for peak 2D angles to directly replicate the methods applied by the previous of both Dingenen and Maykut. We analysed our 2D data independently of the 3D data and did not attempt to determine any time point within the 2D data set from the 3D peak angles. We have made this clearer in the manuscript and this now reads as follows:

"Videos from successful trials were subsequently imported into the Hudl Technique application (Hudl, Agile Sports Technologies Inc., Nebraska, USA) *and analysed independently of the 3D data*."

Statistical analysis:

Can you explain your use of an ICC for construct validity?

Response: The test typically used for validity (Pearson's *r*) is bivariate in design, inappropriate for use with repeated measures of the same variable and cannot account for systematic bias in data sets (George et al [2003], Karras et al [1997]). As a univariate test, an ICC with absolute agreement is therefore a more appropriate choice as it requires a 1:1 ratio to achieve a score reflecting high agreement rather than simply a linear relationship. We have not included this information within the manuscript as we would not consider it of use to the clinical readership but would be happy to do so if the reviewer/editor feel it appropriate.

I am not clear on the inclusion of HIR as a predictor for PKF? Why would HIR influence the KF measurements taken by 2d or 3d?

Response: Our hypothesis was that 3D HIR may potentially explain the likely absence of perfect agreement between 2D and 3D based on the previous work of Ortiz et al. The reviewer is correct though in that this work theorises a potential confounding influence on measurement of HADD rather than KFLEX. However, in investigating HIR as a potential confounding factor, we felt it inappropriate to only do this for one of our chosen variables of interest, hence our application to both.

Results

There is a strong correlation between the KF measures according to Pearson's correlation. How do you explain this?

Response: We feel that this is explained by the bivariate nature of Pearson's which may over-estimate agreement in the presence of linear data behaviour. We have strengthened our discussion of this based on the suggestion of reviewer 2, which reads as follows:

"A further *potential* explanation for this conflict *is* the statistical methodologies employed. Maykut et al (Maykut, et al., 2015) calculated a Pearson's Correlation Coefficient (r) *which, as a bivariate test,* (George, Batterham, & Sullivan, 2003) may over-estimate the agreement between two variables where data demonstrates a linear trend (McGraw & Wong, 1996). *This is reflected by the high (r) produced by the peak KFLEX data from this study (0.74), versus the low (r) produced by the peak HADD data (0.07)."*

Discussion

Can you consider the quality of the camera and distance from the camera along with the software as a possible reason for the results you have found? Can we reasonably expect that someone trying to place on very specific marker on a small screen with their finger to be accurate?

Response:

Line 891: you could have used a marker pen to create markers at the knee and greater trochanter.

Response: The reviewer is correct, we could have, but we chose not to in attempt to be as clinically applicable as possible. We agree though that greater precision is likely to be key for this particular question of 2D versus 3D and make this suggestion in the newly added 'clinical implications' section as suggested by reviewer 2.

Would increased BMI really influence the ability to identify the mid point of the patella and through shorts for the GT?

Response: We agree with the reviewer that an increased BMI is unlikely to affect the visualisation of either the mid-patella or greater trochanter landmarks. We were referring primarily here to the vidualisation of ASIS, which we have added to the manuscript. This now reads as follows:

"This may have negatively affected the accuracy of 2D video digitisation by increasing the visual distortion of necessary bony landmarks given the absence of retroreflective markers, *particularly the ASIS*."

References within text have a number of errors, for example the inclusion of initials or first names.

Response: We apologise for this error, which is linked with how some .RIS files come into Endnote. We have endeavoured to correct this as much as possible.

Highlights

- 2D video analysis using the HUDL application is not a valid method for determining 3D kinematics in runners with PFP.
- Moderate intra-rater reliability for 2D video analysis of kinematics in runners with PFP was established.
- Low (peak hip adduction) to moderate (peak knee flexion) inter-rater reliability for 2D video analysis of kinematics in runners with PFP was established.
- Investigation of other 2D video analysis approaches to increase the precision of clinical gait analysis is warranted.

Is markerless, smart phone recorded two-dimensional video a clinically useful measure of relevant lower limb kinematics in runners with patellofemoral pain? A validity and reliability study

Dr Bradley S Neal [1-2], Dr Simon D Lack [1-2], Dr Christian J Barton [1,3-5], Dr Aleksandra Birn-Jeffrey [1,6], Dr Stuart Miller [1], Dr Dylan Morrissey [1,7]

1. Sports & Exercise Medicine, Queen Mary University of London, United Kingdom.
2. Pure Sports Medicine, London, United Kingdom.
3. La Trobe Sport and Exercise Medicine Research Centre, La Trobe University, Bundoora, Victoria, Australia
4. School of Allied Health, La Trobe University, Melbourne, Victoria, Australia
5. Department of Physical Therapy, University of British Columbia, Vancouver, Canada
6. School of engineering and materials science, Queen Mary University of London, United Kingdom.
7. Physiotherapy Department, Barts Health NHS Trust, London UK

Corresponding Author

Dr Bradley Neal (email: b.s.neal@qmul.ac.uk)

Present address: Sports and Exercise Medicine, Mile End Hospital, Bancroft Road, London E1 4DG, United Kingdom.

Abstract word count: 199

Word count: 3879 (excluding tables and figures)

Is markerless, smart phone recorded two-dimensional video a clinically useful measure of relevant lower limb kinematics in runners with patellofemoral pain? A validity and reliability study

Abstract

Objectives: Investigate the validity and reliability of markerless, smart phone collected, two-dimensional (2D) video, analysed using the 'Hudl technique' application, compared to three-dimensional (3D) kinematics during running, in participants with patellofemoral pain (PFP).

Design: Validity/reliability study

Setting: Biomechanics laboratory

Participants: Males/females with PFP (n=21, 10 males, 11 females, age 32.1 months [±12.9]).

Main Outcome Measures: Manually synchronised 2D and 3D measurement of peak hip adduction (HADD) and peak knee flexion (KFLEX) during running.

Results: 2D and 3D measures of peak KFLEX (p=0.02, *d*=1.13), but not peak HADD (p=0.25, *d*=-0.27), differed significantly. Poor validity was identified for 2D measurement of peak HADD (ICC 0.06, 95% CI -0.35, 0.47) and peak KFLEX (ICC 0.42, 95% CI (-0.10, 0.75). Moderate intra-rater reliability was identified for both variables (ICC 0.61-65), alongside moderate inter-rater reliability for peak KFLEX (ICC 0.71) and poor inter-rater reliability for peak HADD (ICC 0.31).

Conclusions: Measurement of peak HADD and KFLEX in runners with PFP using markerless, smart phone collected 2D video, analysed using the Hudl Technique Application is invalid, with poor to moderate reliability. Investigation of alternate 2D video approaches to increase precision is warranted. At present, 2D video analysis of running using Hudl Technique cannot be advocated.

Key Words

Patellofemoral Pain, Running, Kinematics, Validity

INTRODUCTION

Recreational running is a common form of exercise (Linton & Valentin, 2018) associated with both positive health benefits (Lee, et al., 2017) and high rates of musculoskeletal injury (19-94%). (Saragiotto, et al., 2014) The knee is reported to be the most prevalent joint involved in running-related musculoskeletal injury, (Linton & Valentin, 2018; Taunton, et al., 2002) with patellofemoral pain (PFP) the most prevalent diagnosis (17%). (Taunton, et al., 2002) New incident cases of PFP amongst recreational runners were recently reported to be 6%. (Bradley S Neal, Lack, et al., 2018)

Whilst musculoskeletal injuries are multi-factorial, (Bittencourt, et al., 2016) peak hip adduction (HADD) during running has been reported as a risk factor for future PFP development in female runners (Noehren, Hamill, & Davis, 2013) and is associated with the persistence of PFP in mixed-sex cohorts. (B. S. Neal, Barton, Gallie, O'Halloran, & Morrissey, 2016) Peak HADD of ≥20˚ and a reduction in peak HADD of 5˚ are also reported to be a potential treatment target and mechanism of effect underpinning running retraining in PFP respectively. (Bradley S Neal, Barton, Birn-Jeffrey, Daley, & Morrissey, 2018; Noehren, Scholz, & Davis, 2011; R.W. Willy, Scholz, & Davis, 2012) Peak knee flexion is also a variable of interest in runners with PFP. It is reported to correlate with patellofemoral joint stress (Lenhart, Thelen, Wille, Chumanov, & Heiderscheit, 2014) and is also associated with kinesiophobia, with females with PFP demonstrating lower peak knee flexion angles during stair descent. (de Oliveira Silva, Barton, Pazzinatto, Briani, & de Azevedo, 2016) Altering peak knee flexion may be associated with symptomatic improvements after a step rate retraining intervention. (Bradley S Neal, Barton, et al., 2018)

Guidelines for the measurement of these running kinematics with two-dimensional (2D) video in clinical practice are already in place. (Souza, 2016) Two previous studies have reported concurrent validity and reliability of high frame rate 2D video in comparison to three-dimensional (3D) kinematic motion capture, for measuring peak HADD. (Bart Dingenen, et al., 2017; Maykut, Taylor-Haas, Paterno, DiCesare, & Ford,

2015) Maykut et al. reported a significant, moderate correlation between 2D and 3D measurement for peak HADD during treadmill running (r=0.53-0.62). (Maykut, et al., 2015) In addition, Dinengen et al. reported a significant, positive correlation for peak HADD during over ground running, using a discrete 2D variable to predict an entire 3D kinematic curve from initial ground contact through to toe off. (Bart Dingenen, et al., 2017) Whilst both of these studies reported their methods to be reliable (ICC 0.90-0.99), given that runners with PFP demonstrate differing kinematics compared to matched controls, (B. S. Neal, et al., 2016) the investigation of asymptomatic runners limits the external applicability to clinical populations. Furthermore, both studies analysed their 2D videos using Dartfish software, which may be prohibitive in clinical practice due to high costs and complexity of use relative to simpler, mobile phone-based applications.

Previous studies investigating construct validity for 2D video to measure peak HADD have not identified optimal agreement between 2D and 3D measurement. Ortiz et al hypothesised that transverse plane hip motion may affect the accuracy of 2D measurement of frontal plane hip kinematics during a jump/land task. (Ortiz, et al., 2016) Runners with PFP have also been reported to demonstrate increased peak hip internal rotation (HIR) in comparison to controls. (Noehren, Pohl, Sanchez, Cunningham, & Lattermann, 2012; Noehren, Sanchez, Cunningham, & McKeon, 2012; Souza & Powers, 2009a, 2009b; R. W. Willy, Manal, Witvrouw, & Davis, 2012) Transverse plane motion at the hip is coupled with HADD and tibial abduction, referred to in combination as dynamic knee valgus. (Powers, 2010) Determining the impact of this movement direction on the variability observed between 2D and 3D measurement may provide insight into the source of previously reported sub-optimal agreement. (Ortiz, et al., 2016)

This study aimed to determine whether clinicians can use a simple, readily available tool to measure important lower limb kinematic variables during running. The primary objective was to investigate the concurrent validity and intra- and inter-rater reliability of markerless, high frame rate 2D video, recorded using a smart phone, with reference to 3D kinematic motion capture. The null hypothesis was that smart phone collected 2D video would not give useful measurements of acceptable accuracy with respect to 3D kinematic analyses and as such, a secondary objective was to investigate the source of any identified disagreement.

METHODS

The Queen Mary Ethics of Research Committee (QMREC2014/24/103) gave ethical approval for this study.

Sample size calculation

Using 2D and 3D peak HADD means and a pooled SD from previous work (2D HADD 11.2° [±2.7], 3D HADD 14.0° [3.7]) (Maykut, et al., 2015) and equations for dependent samples t-tests, 21 participants were required to achieve α 5% and β 80% (calculated using G*Power 3.1.9.2, Heinrich-Heine University, Germany). 21 participants with PFP (10 male, 11 female) were conveniently sampled from local sports medicine clinics (see table 1). All participants provided written informed consent prior to participating.

Table 1: Participant characteristics

Variable	**Mean (SD)**
Age (years)	32.1 (12.9)
Height (cm)	169.1 (45.2)
Mass (kg)	69.8 (19.6)
BMI	23.2 (2.6)
Tegner scale	5.5 (1.3)
Symptom duration (months)	53.1 (± 84.5)
Kujala scale	76.2 (± 12.9)
Average NRS	4.7 (± 2.0)

Key: SD=standard deviation; cm=centimeters; kg=kilograms; BMI=body mass index; NRS=numerical rating scale.

Participants

To be eligible, participants were required to have insidious onset retropatellar or peripatellar pain for a minimum of one month, during at least one activity including running, squatting, stair ambulation and jumping. (Crossley, et al., 2016) The Tegner Activity Scale was collected to act as a constant measure across a heterogeneous cohort of participants with PFP who participated in a variety of sports and hobbies.

(Lysholm & Tegner, 2007) Participants below the age of 18 or over the age of 50, or those with traumatic symptoms, patellofemoral instability, tibiofemoral pathology or other concomitant pathology were excluded. Height and mass were collected to allow for the calculation of BMI, reported to be higher in those with persistent PFP. (Hart, Barton, Khan, Riel, & Crossley, 2017) Symptom duration, Kujala scale and average pain over the last 3 months using a numerical rating scale between 0 and 10 (0 = no pain and 10 = worst pain imaginable) were collected as a reflection of symptom severity and persistence, reported to alter running kinematics. (Fox, Ferber, Saunders, Osis, & Bonacci, 2018)

3D kinematics

Kinematic data were collected during running using a four-camera, infrared motion analysis system using Odin software (CX-1, Codamotion, Charnwood Dynamics Limited, Leicestershire, UK), sampling at 200Hz. (Lack, et al., 2014) 24 infrared markers; eight individual markers (10mm) and four rigid clusters of four markers (140mm), were placed adhering to the CAST protocol. (Cappello, Cappozzo, La Palombara, Lucchetti, & Leardini, 1997) Individual markers were placed on the anterior superior iliac spine (ASIS), posterior superior iliac spine (PSIS), lateral calcaneal process and head of 5th metatarsal, with rigid clusters placed on the mid-point of each thigh and shank segment. Foot markers were placed on the participants shoe as an estimation of the anatomical location, given the potential for barefoot running to effect running kinematics. (Hall, Barton, Jones, & Morrissey, 2013) Unpublished intra-rater reliability of kinematic marker placement for the primary investigator (BN) has previously been found to be moderate to excellent (ICC 0.62 – 0.93).

Rigid clusters were secured using a combination of adjustable elastic straps and cohesive self-adherent bandage, with individual markers applied using double-sided adhesive tape and secured with transparent surgical tape. Virtual markers were also identified on the femoral epicondyles and the ankle malleoli, to allow for the calculation of relevant joint centres. The knee joint centre was estimated as the mid-point between the femoral epicondyle markers and the hip joint centre was estimated as a projection within the pelvis frame using previously described methods. (Bell, Pedersen, & Brand, 1990) Joint centre calculation did not differ between male and female participants.

2D kinematics

2D kinematic data were captured using two high frame-rate smartphone cameras (iPhone 6, Apple Corporation, California, USA) recording at 240/frames per second. Cameras were mounted on stable tripods 1.0 metre from the laboratory floor. The camera recording in the sagittal plane was placed at a distance of 2.5 metres from the centre of the ground-embedded force plate (type 9281CA, Kistler Corporation, Switzerland), which participants subsequently ran past. The camera recording in the frontal plane was placed 6.5 metres from the centre of the ground-embedded force plate, which participants ran directly towards (see figure 1).

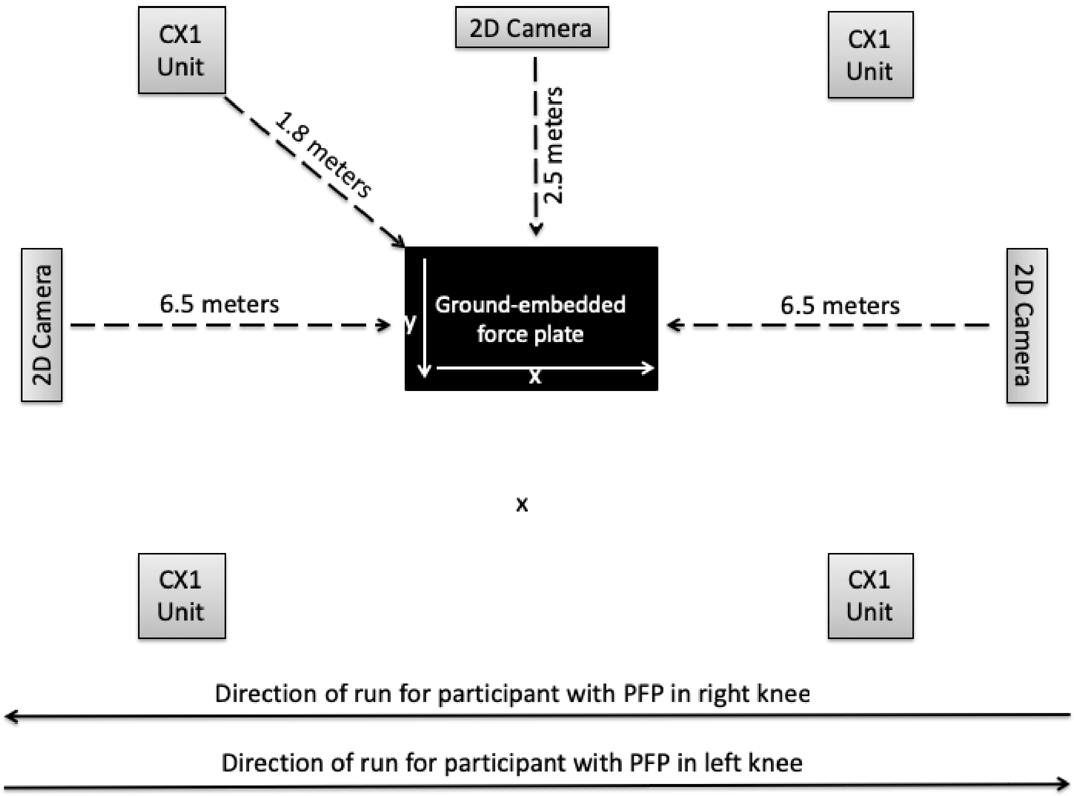

Figure 1: human performance laboratory set up detailing the location of 3D and 2D cameras

Experimental protocol

Both 2D and 3D data were captured during trials of over-ground running in a human performance laboratory. Participants were provided with neutral running shoes in their required size (Asics Nimbus, Asics, Cheshire, UK), to minimise potential effects of footwear variation on running kinematics. (Hall, et al., 2013) Participants were instructed to run in a straight line for a distance of approximately 10.0 meters at a self-selected speed, landing the foot of their symptomatic limb on the ground-embedded force plate, sampling at 1000Hz. The ground-embedded force plate was 5.0 metres from the trial start-point, with participants typically making contact with their fifth step as they ran through. Several practice runs were permitted to allow for familiarisation and to ensure adequate force plate contact during a participant's natural running gait without deceleration. This process was repeated until five successful trials were obtained, with a successful trial defined as an appropriate landing of the correct foot directly onto the force plate without obvious adjustment of running gait. Each trial was initiated by verbal countdown by a member of the research team, with the 3D system and both 2D cameras manually synchronised using a numerical countdown.

Data analysis

To reduce the potential for type I error, data pertaining to one limb only were entered into the analysis. (Menz, 2005) For participants with bilateral symptoms, the limb that rated the highest on the numerical rating scale was evaluated. In the presence of equivalent symptoms the dominant limb was evaluated, defined by the limb that the participant would use to kick a ball.

3D kinematic analysis

Data were analysed offline using a customised Matlab program (version 2015, Mathworks, Natick, Massachussets, USA). A 20*N* threshold from the ground-embedded force plate was used to determine initial contact and toe-off respectively. Kinematic data were processed within this event window, defined as running stance phase. An international society of biomechanics advocated XZY (sagittal, frontal, transverse) cardan rotation sequence was used. Peak joint angles for both peak hip adduction (HADD) and knee flexion (KFLEX) were visualised and subsequently exported to a Microsoft Excel (Microsoft Corporation, Albuquerque, New Mexico, USA) for statistical analysis.

2D kinematic analysis

Videos from successful trials were subsequently imported into the Hudl Technique application (Hudl, Agile Sports Technologies Inc., Nebraska, USA) and analysed independently of the 3D data. 2D data analysis was completed using a tablet device with a 25.9cm screen (5th generation iPad, Apple Corporation, California, USA). Two independent 2D angles, hip adduction (HADD) and knee flexion (KFLEX) were identified using the angle tool. Use of the zoom function within the Hudl technique application was permitted at the discretion of the analyser, to ensure optimal visualization of the relevant anatomical landmarks.

HADD was determined using methods described by Dingenen et al, where the contralateral pelvic drop (CLPD) angle is added from the femoral adduction (FADD) angle. (Bart Dingenen, et al., 2017) CLPD angle was defined as the angle formed by a horizontal line from the stance limb anterior superior iliac spine (ASIS) (referenced from the laboratory floor) and the swing limb ASIS (see figure 2). FADD angle was defined as the angle formed by a horizontal line from the stance limb ASIS (referenced from the laboratory floor) and the centre of the stance limb tibiofemoral joint (an estimation of the knee joint centre) (see figure 2). Within the Hudl technique application, the tool reflects an angle relative to 90˚ and the FADD angle was therefore determined by subtracting the angle produced by the tool from 90˚.

KFLEX was defined as the angle formed by a line drawn from the stance limb greater trochanter to the lateral femoral condyle and a second line drawn from the stance limb lateral femoral condyle to the stance limb lateral malleolus (see figure 2). Within the Hudl technique application, a vertical line in the sagittal plane is reflective of 180˚ and the KFLEX angle was therefore determined by subtracting the angle produced by the tool from 180˚. For both variables, a peak angle was estimated, determined to be when the participant reached the peak of mid-stance, manually defined as the point where maximal foot contact had occurred and no upward/downward motion was occurring. (Maykut, et al., 2015)

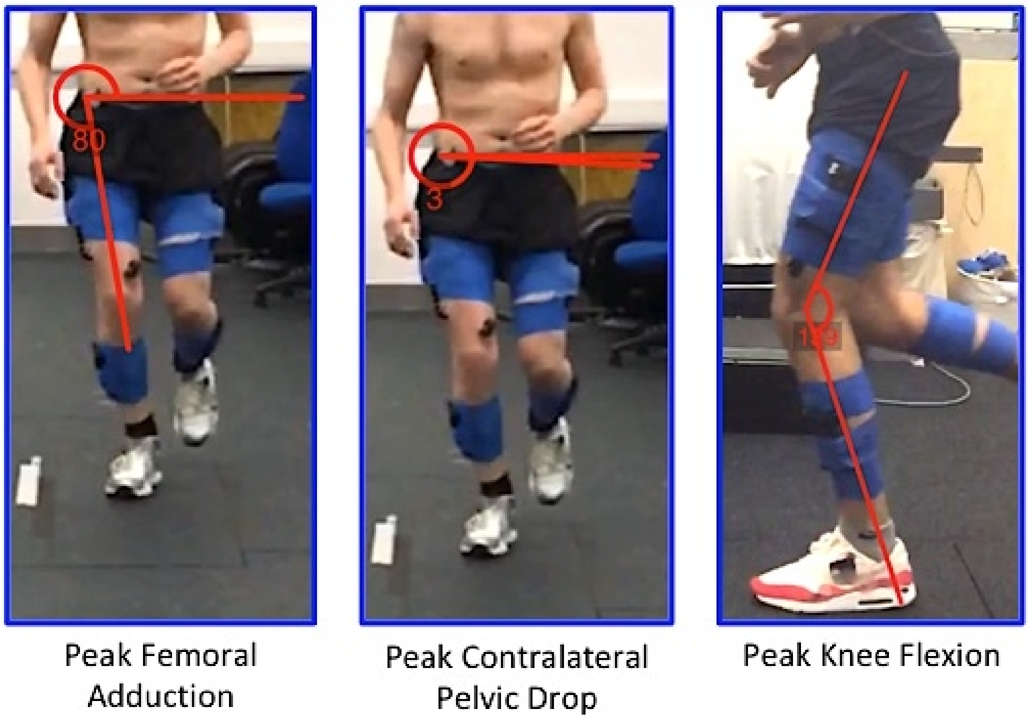

Figure 2: determination of 2D joint angles at the hip and knee

<u>Statistical analysis</u>

All analyses were performed using SPSS (version 22 for MacOS, IBM, New York, USA). The mean of the five 2D and 3D trials were calculated for each participant for both variables of interest (HADD and KFLEX). The difference between the 2D and 3D means was determined using two-tailed, dependent samples t-tests. Single measure ICCs with 95% confidence intervals were calculated using a two-way mixed effects model with absolute agreement, to determine construct validity and both intra- and inter-rater reliability. ICCs were defined as excellent (> 0.90), good (0.75-0.90), moderate (0.50-0.75) and poor (< 0.50) respectively (Koo & Li, 2016). A standard error of measure (SEM) was also calculated to allow for clinical interpretation of reliability data. Bland and Altman plots with 95% limits of agreement (LOA) were used to visually represent the agreement between the 2D and 3D values (Bland & Altman, 1986). Scatter plots were used to visualise the directionality of the relationship between 2D and 3D measurement, with a Pearson's correlation coefficient (*r*) also calculated to allow for comparisons with previous work.

2D peak HADD and KFLEX values from the first run trial of all participants were analysed twice by the primary investigator (BN), with 24 hours between analyses, to determine intra-rater reliability. 2D peak HADD and KFLEX values from the first run trial of all participants were also analysed by a second investigator (SL) and compared to the initial analyses of the primary investigator (BN), to determine inter-rater reliability. Finally, to assess the influence of including 3D peak hip internal rotation (HIR) in a predictive model, a backward linear regression was performed, with the F change statistic used to determine if 3D HIR explains the hypothesised imperfect agreement between 2D and 3D measurement.

RESULTS

Construct validity

There was a significant difference between 2D and 3D measured peak KFLEX, whereas peak HADD was not significantly different between 2D and 3D measures (table 2). ICCs identified a poor correlation for both peak HADD and peak KFLEX between 3D and 2D measurement (table 3).

Table 2: 3D and 2D data for both variables

Variable	**3D Measurement (Mean ± SD)**	**2D Measurement (Mean ± SD)**	**Difference (Mean ± SD)**	***P***	***d***
HADD	12° ± 4.7	13° ± 3.2	-1°	0.25	-0.27
KFLEX	38° ± 5.5	43° ± 3.3	-5°	<0.01*	-1.13

Key: 3D= three dimensional; 2D=two dimensional; SD=standard deviation HADD=hip adduction; KFLEX=knee flexion.

Table 3: construct validity data for both variables comparing 3D and 2D measurement

Outcome	**HADD**	**KFLEX**
ICC (95% CI)	0.06 (-0.35, 0.47)	0.42 (-0.10, 0.75)
Upper LOA	10.9	7.4
Lower LOA	-10.9	-7.4

Key: HADD=hip adduction; KFLEX=knee flexion; SD=standard deviation; ICC=intraclass correlation coefficient; CI=confidence interval; LOA=limits of agreement.

A multiple variable, backward linear regression was calculated to predict 3D peak HADD (dependent variable) using 2D HADD (independent variable$_1$) and 3D HIR (independent variable$_2$). R^2 of the model was 0.06, with a non-significant F change (0.07, p=0.93) identified after the removal of 3D HIR (R^2 change -0.01).

A second multivariable backward linear regression was calculated to predict 3D KFLEX (dependent variable) using both 2D KFLEX (independent variable$_1$) and 3D HIR (independent variable$_2$). R^2 of the model was 0.60, with a non-significant F change (3.76, p=0.06) identified after the removal of 3D HIR (R^2 change -0.08).

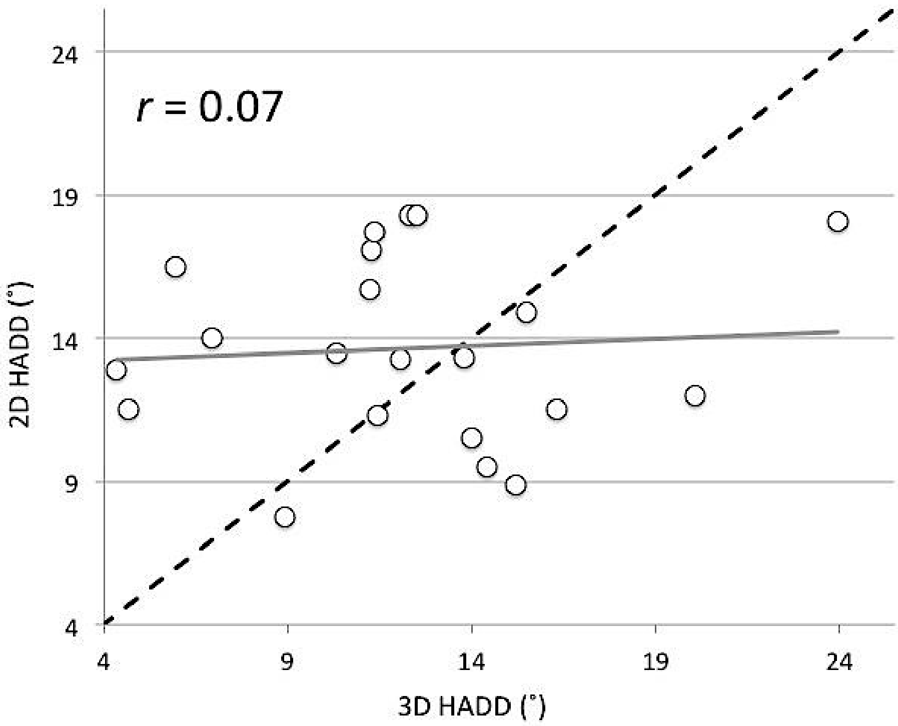

Figure 3: scatter plot for peak 3D and 2D HADD
Key: dashed line represents a line of identity; solid line represents a line of best fit

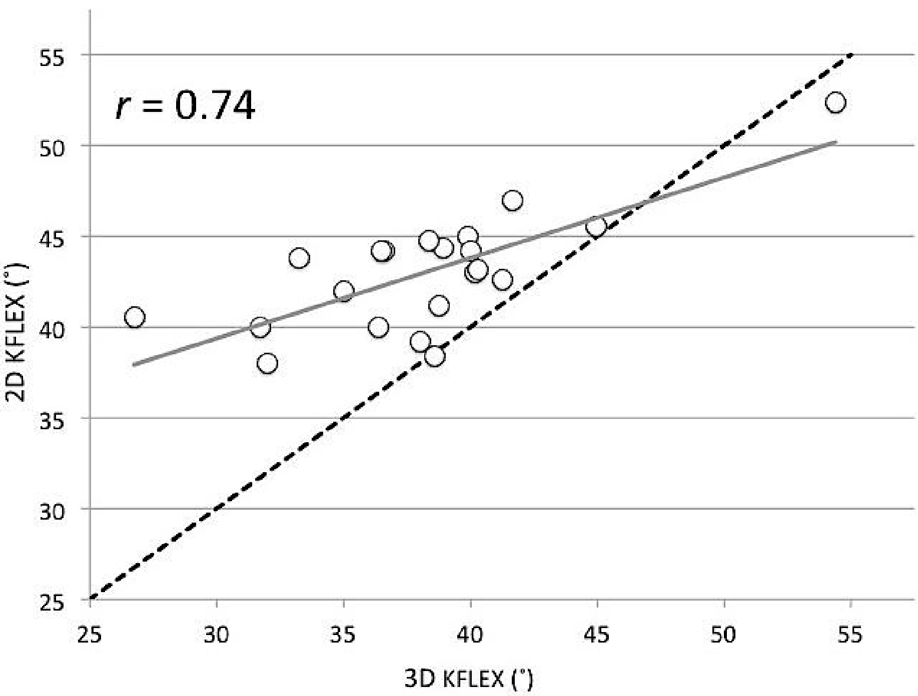

Figure 4: scatter plot for peak 3D and 2D KFLEX
Key: dashed line represents a line of identity; solid line represents a line of best fit

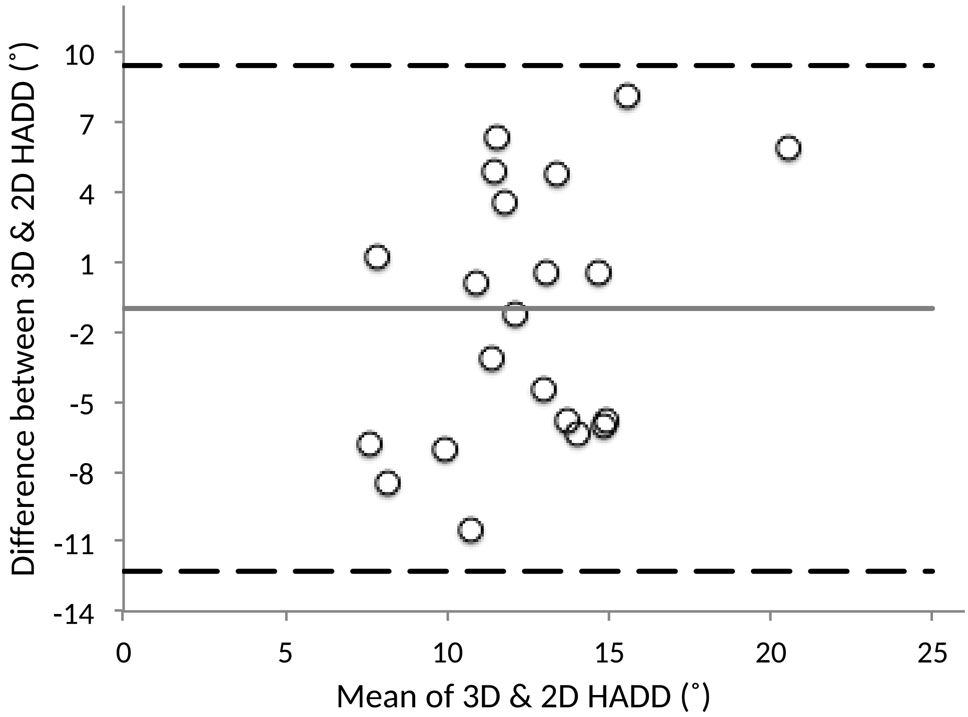

Figure 5: Bland and Altman plot for peak HADD
Key: dashed lines represent upper and lower limits of agreement, solid line represents the pooled mean difference between 3D and 2D measurement.

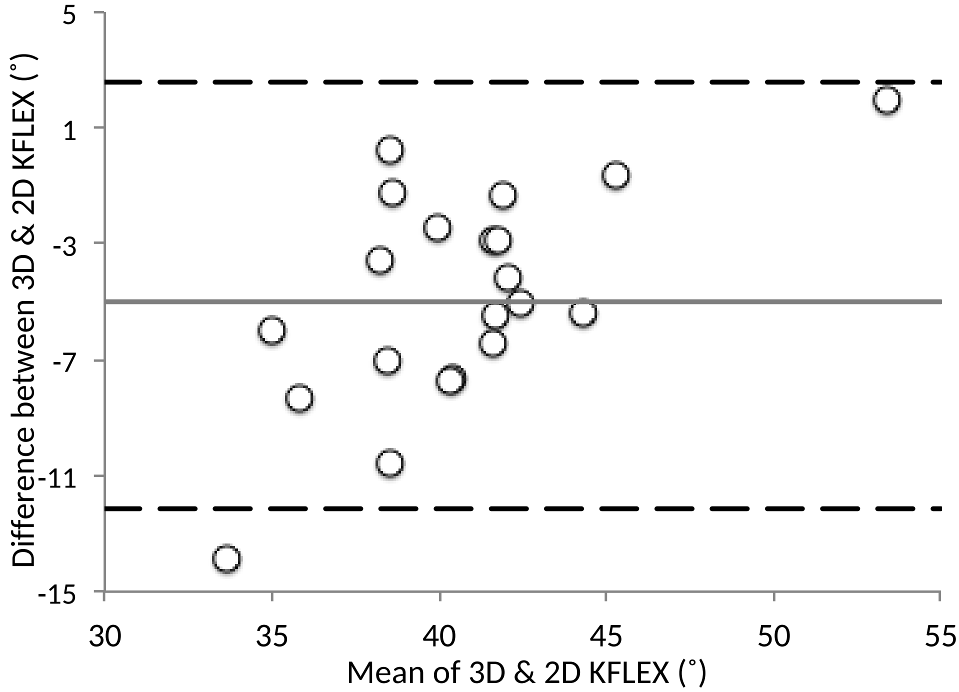

Figure 6: Bland and Altman plot for peak KFLEX
Key: dashed lines represent upper and lower limits of agreement, solid line represents the pooled mean difference between 3D and 2D measurement.

Intra-rater reliability

Moderate intra-rater reliability was identified for peak HADD (ICC 0.65 95% CI 0.34, 0.83, SEM 1.8˚) and peak KFLEX (ICC 0.61 95% CI -0.09, 0.87, SEM 2.7˚).

Inter-rater reliability

Poor inter-rater reliability was identified for peak HADD (ICC 0.31 95% CI -0.06, 0.64, SEM 3.1˚). Moderate inter-rater reliability was identified for peak KFLEX (ICC 0.71 95% CI 0.16, 0.89, SEM 1.4˚).

DISCUSSION

Accepting our null hypothesis, 2D measurement of both peak HADD and peak KFLEX was shown to be invalid and have poor to moderate reliability, reflected by low ICCs and wide limits of agreement. These data suggest that markerless, mobile phone collected 2D video, analysed using the Hudl Technique application, does not have acceptable accuracy to quantify either peak HADD or KFLEX during over ground running in individuals with PFP.

Our validity data for peak HADD conflict with the work of both Maykut et al (Maykut, et al., 2015) and Dingenen et al (Bart Dingenen, et al., 2017), who reported significant correlations between 2D and 3D measured peak HADD, despite recording their 2D video at a lower collection frequency. The primary explanation for this disagreement may be the software used to assess the 2D videos. We evaluated the construct validity of the Hudl Technique application, given its ease of clinical application. Hudl Technique is free of cost at the point of access and can be installed on a variety of devices (mobile phones and tablets) and operating systems. The Dartfish software (Dartfish, Fribourg, Switzerland) used in previous studies may offer greater precision, where digitizing 2D video is completed using a mouse on a larger screen, rather than the assessor's finger on a smaller touch screen. The limitation of Dartfish as a method of 2D video analysis is the associated cost (£204-£880 per calendar year).

An additional discrepancy between our study and the work of both Maykut et al (Maykut, et al., 2015) and Dingenen et al (Bart Dingenen, et al., 2017) is the investigation of a cohort of participants with PFP in comparison to asymptomatic participants. Reflective of a typical cohort with persistent PFP (mean symptom duration 53.1 months), our participants had a higher BMI (mean 23.2) than the previously studied asymptomatic cohorts. This may have negatively affected the accuracy of 2D video digitisation by increasing the visual distortion of necessary bony landmarks given the absence of retroreflective markers, particularly the ASIS. Furthermore, our PFP cohort had a lower physical activity level (mean Tegner Scale

5.5) in comparison to the elite asymptomatic cohorts investigated by both Maykut et al (Maykut, et al., 2015) and Dingenen et al (Bart Dingenen, et al., 2017) (estimated Tegner Scale 8-10). Elite runners are reported to have more consistent kinematics than recreational runners (Clermont, Osis, Phinyomark, & Ferber, 2017), which is likely to have resulted in a more stable mean and thus, increased agreement between 2D and 3D measurement (Bart Dingenen, Barton, Janssen, Benoit, & Malliaras, 2018).

A further potential explanation for this conflict is the statistical methodologies employed. Maykut et al (Maykut, et al., 2015) calculated a Pearson's Correlation Coefficient (*r*) which, as a bivariate test, (George, Batterham, & Sullivan, 2003) may over-estimate the agreement between two variables where data demonstrates a linear trend (McGraw & Wong, 1996). This is reflected by the high (*r*) produced by the peak KFLEX data from this study (0.74), versus the low (*r*) produced by the peak HADD data (0.07). Dingenen et al employed statistical parametric mapping (Bart Dingenen, et al., 2017), which does not confirm that the 2D method used can accurately predict a discrete 3D value at a specific point within the gait cycle. Clinicians often seek a discrete kinematic variable within the gait cycle to employ clinical prediction rules, such as a 5° reduction in peak HADD as a predictor for running retraining success, (Noehren, et al., 2011; R.W. Willy, et al., 2012) thus limiting the clinical applicability of these data. A summary of the discrepancies between this study and previous work is presented in table 4.

Table 4: methodological comparison between studies.

	This study	**(Maykut, et al., 2015)**	**(Bart Dingenen, et al., 2017)**
Population	Physically active persons with PFP	Asymptomatic elite runners	Asymptomatic elite athletes
Tegner scale	5	9 (estimated)	9 (estimated)
Mean BMI	23.2	20.0	21.1
Mean age (years)	32.1	19.9	18.7
Running method	Over ground	Treadmill	Over ground
2D video recording frequency	240 fps	60 fps	50 fps
Retroreflective markers	No	Yes	Yes
2D analysis software	Hudl Technique	Dartfish	Dartfish
Statistical method	ICC	Pearson's *r*	SPM
Analysis screen size	Tablet	Computer	Computer
Frontal plane camera distance from axis	6.5 meters	?	4.5 meters

Key: PFP=patellofemoral pain; ?=unable to determine; fps=frames per second; SPM=statistical parametric mapping. Grey shading indicates commonalities between studies.

Our novel investigation of peak KFLEX also demonstrates a poor agreement between 2D video and 3D kinematic motion capture. There is a linear pattern to these data, which results in a Pearson's *r* that over-estimates construct validity (r=0.74 versus ICC=0.42). There also appears to be a systematic bias within these data, with 2D video consistently over-predicting peak KFLEX by a mean of 5˚. Ortiz et al (219) hypothesised that transverse plane hip motion may affect the accuracy of 2D measured running kinematics. Consistent with this hypothesis, there is a statistical trend towards 3D peak HIR being a covariate for this outcome (F change 3.76, p=0.06, R2 change -0.08). Whilst this may explain the systematic bias within these data, this potential model has limited clinical applicability, as transverse plane hip data are not collectable using 2D cameras.

Limitations and future directions

This study is not without limitations, which must be considered when interpreting the results. In an attempt to best replicate clinical practice, participants completed only a short over ground run, with data collected on the fifth step on average. Dingenen et al (Bart Dingenen, et al., 2018) recently reported that a minimum of seven steps are required to allow for a stable mean of a 2D measured kinematic variable. These data refer to analysis completed with Kinovea (http://www.kinovea.org), software that is free of cost at the point of access to Microsoft Windows users. Kinovea offers comparable analysis precision to Dartfish and has been reported to be both inter- and intra-rater reliable for measuring a variety of 2D running kinematic variables (Bart Dingenen, et al., 2018) when data were collected using retroflective markers. Given the apparent potential for increased precision to result in greater construct validity, a future study using either Dartfish or Kinovea involving runners with PFP is warranted.

Only two kinematic variables were assessed in this study and it may be that other kinematic variables prove to be both valid and reliable if investigated by future studies. It could also be that repeating this study using a treadmill running protocol similar to that used by Maykut et al (Maykut, et al., 2015) may return a different outcome. Kinematic comparisons between treadmill and over ground running have been reported to be equivalent (Sinclair, et al., 2013) and a treadmill protocol would allow for the frontal plane camera to be placed closer to the runner, increasing 2D video quality and reducing the potential for parallax error. (B. Dingenen, et al., 2019) Finally, a single video, rather than mean pooled data, were used for the investigation of reliability, differing from the investigation of validity. Whilst this decision was made apriori, analysis of mean pooled data may have yielded different reliability results.

Clinical implications

Whilst the results of this study suggest that markerless, smart phone collected high frame rate 2D video analysed using the Hudl technique application is invalid, there are some implications for clinical practice. Rather than being concerned about maximising video frame rate, attention should be given to placing the 2D camera(s) as close to the runner as possible, to increase quality and reduce parallax error potential. This is most easily achieved using a treadmill rather than over ground running. In addition, use of retroreflective markers is encouraged to maximize ease of identifying relevant bony landmarks, especially those that may be obscured by adipose tissue or clothing. Finally, clinicians are encouraged to analyse 2D data using a large screen and with software that allows for increased precision via use of a computer mouse (or equivalent), rather than a smaller tablet with a touch screen, which is likely to yield inaccurate results.

CONCLUSION

Measurement of both peak HADD and KFLEX in runners with PFP using mobile phone collected, high frame rate 2D video, analysed using the Hudl Technique Application is invalid, with poor to moderate reliability. This may be attributed to the employed 2D video or statistical methodologies, but could also be explained by the increased variability in running kinematics of runners with PFP. Further investigation of methodologies with increased precision is warranted, aiming to improve the ability of high frame rate 2D video to accurately predict 3D kinematics in the clinical setting. At present, clinical gait analysis conducted using the Hudl Technique application should be interpreted with caution, as the validity or reliability of 2D measurement cannot be guaranteed.

Reference List

Bell, A. L., Pedersen, D. R., & Brand, R. A. (1990). A comparison of the accuracy of several hip center location prediction methods. *J Biomech, 23*, 617-621.

Bittencourt, N., Meeuwisse, W., Mendonça, L., Nettel-Aguirre, A., Ocarino, J., & Fonseca, S. (2016). Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *Br J Sports Med*, bjsports-2015-095850.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*, 307-310.

Cappello, A., Cappozzo, A., La Palombara, P. F., Lucchetti, L., & Leardini, A. (1997). Multiple anatomical landmark calibration for optimal bone pose estimation. *Human movement science, 16*, 259-274.

Clermont, C. A., Osis, S. T., Phinyomark, A., & Ferber, R. (2017). Kinematic Gait Patterns in Competitive and Recreational Runners. *J Appl Biomech, 33*, 268-276.

Crossley, K. M., Stefanik, J. J., Selfe, J., Collins, N. J., Davis, I. S., Powers, C. M., McConnell, J., Vicenzino, B., Bazett-Jones, D. M., Esculier, J. F., Morrissey, D., & Callaghan, M. J. (2016). 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 1: Terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome measures. *Br J Sports Med, 50*, 839-843.

de Oliveira Silva, D., Barton, C. J., Pazzinatto, M. F., Briani, R. V., & de Azevedo, F. M. (2016). Proximal mechanics during stair ascent are more discriminate of females with patellofemoral pain than distal mechanics. *Clin Biomech (Bristol, Avon), 35*, 56-61.

Dingenen, B., Barton, C., Janssen, T., Benoit, A., & Malliaras, P. (2018). Test-retest reliability of two-dimensional video analysis during running. *Physical therapy in Sport, 33*, 40-47.

Dingenen, B., Malliaras, P., Janssen, T., Ceyssens, L., Vanelderen, R., & Barton, C. J. (2019). Two-dimensional video analysis can discriminate differences in running kinematics between recreational runners with and without running-related knee injury. *Phys Ther Sport, 38*, 184-191.

Dingenen, B., Staes, F. F., Santermans, L., Steurs, L., Eerdekens, M., Geentjens, J., Peers, K. H., Thysen, M., & Deschamps, K. (2017). Are two-dimensional measured frontal plane angles related to three-dimensional measured kinematic profiles during running? *Physical therapy in Sport*.

Fox, A., Ferber, R., Saunders, N., Osis, S., & Bonacci, J. (2018). Gait Kinematics in Individuals with Acute and Chronic Patellofemoral Pain. *Med Sci Sports Exerc, 50*, 502-509.

George, K., Batterham, A., & Sullivan, I. (2003). Validity in clinical research: a review of basic concepts and definitions☆. *Physical therapy in Sport, 4*, 115-121.

Hall, J. P., Barton, C., Jones, P. R., & Morrissey, D. (2013). The biomechanical differences between barefoot and shod distance running: a systematic review and preliminary meta-analysis. *Sports Med, 43*, 1335-1353.

Hart, H. F., Barton, C. J., Khan, K. M., Riel, H., & Crossley, K. M. (2017). Is body mass index associated with patellofemoral pain and patellofemoral osteoarthritis? A systematic review and meta-regression and analysis. *Br J Sports Med, 51*, 781-790.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med, 15*, 155-163.

Lack, S., Barton, C., Malliaras, P., Twycross-Lewis, R., Woledge, R., & Morrissey, D. (2014). The effect of anti-pronation foot orthoses on hip and knee kinematics and muscle activity during a functional step-up task in healthy individuals: a laboratory study. *Clin Biomech 29*, 177-182.

Lee, D. C., Brellenthin, A. G., Thompson, P. D., Sui, X., Lee, I. M., & Lavie, C. J. (2017). Running as a Key Lifestyle Medicine for Longevity. *Prog Cardiovasc Dis, 60*, 45-55.

Lenhart, R. L., Thelen, D. G., Wille, C. M., Chumanov, E. S., & Heiderscheit, B. C. (2014). Increasing running step rate reduces patellofemoral joint forces. *Med Sci Sports Exerc, 46*, 557-564.

Linton, L., & Valentin, S. (2018). Running with injury: A study of UK novice and recreational runners and factors associated with running related injury. *J Sci Med Sport*.

Lysholm, J., & Tegner, Y. (2007). Knee injury rating scales. *Acta Orthopaedica, 78*, 445-453.

Maykut, J. N., Taylor-Haas, J. A., Paterno, M. V., DiCesare, C. A., & Ford, K. R. (2015). Concurrent validity and reliability of 2d kinematic analysis of frontal plane motion during running. *Int J Sports Phys Ther, 10*, 136-146.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods, 1*, 30.

Menz, H. B. (2005). Analysis of paired data in physical therapy research: time to stop double-dipping? *J Orthop Sports Phys Ther, 35*, 477-478.

Neal, B. S., Barton, C. J., Birn-Jeffrey, A., Daley, M., & Morrissey, D. (2018). The effects & mechanisms of increasing running step rate: A feasibility study in a mixed-sex group of runners with patellofemoral pain. *Physical therapy in Sport*.

Neal, B. S., Barton, C. J., Gallie, R., O'Halloran, P., & Morrissey, D. (2016). Runners with patellofemoral pain have altered biomechanics which targeted interventions can modify: A systematic review and meta-analysis. *Gait Posture, 45*, 69-82.

Neal, B. S., Lack, S. D., Lankhorst, N. E., Raye, A., Morrissey, D., & van Middelkoop, M. (2018). Risk factors for patellofemoral pain: a systematic review and meta-analysis. *Br J Sports Med*, bjsports-2017-098890.

Noehren, B., Hamill, J., & Davis, I. (2013). Prospective evidence for a hip etiology in patellofemoral pain. *Med Sci Sports Exerc, 45*, 1120-1124.

Noehren, B., Pohl, M. B., Sanchez, Z., Cunningham, T., & Lattermann, C. (2012). Proximal and distal kinematics in female runners with patellofemoral pain. *Clin Biomech 27*, 366-371.

Noehren, B., Sanchez, Z., Cunningham, T., & McKeon, P. O. (2012). The effect of pain on hip and knee kinematics during running in females with chronic patellofemoral pain. *Gait Posture, 36*, 596-599.

Noehren, B., Scholz, J., & Davis, I. (2011). The effect of real-time gait retraining on hip kinematics, pain and function in subjects with patellofemoral pain syndrome. *Br J Sports Med, 45*, 691-696.

Ortiz, A., Rosario-Canales, M., Rodriguez, A., Seda, A., Figueroa, C., & Venegas-Rios, H. L. (2016). Reliability and concurrent validity between two-dimensional and three-dimensional evaluations of knee valgus during drop jumps. *Open Access J Sports Med, 7*, 65-73.

Powers, C. M. (2010). The influence of abnormal hip mechanics on knee injury: a biomechanical perspective. *J Orthop Sports Phys Ther, 40*, 42-51.

Saragiotto, B. T., Yamato, T. P., Hespanhol Junior, L. C., Rainbow, M. J., Davis, I. S., & Lopes, A. D. (2014). What are the main risk factors for running-related injuries? *Sports Med, 44*, 1153-1163.

Sinclair, J., Richards, J., Taylor, P. J., Edmundson, C. J., Brooks, D., & Hobbs, S. J. (2013). Three-dimensional kinematic comparison of treadmill and overground running. *Sports Biomech, 12*, 272-282.

Souza, R. B. (2016). An Evidence-Based Videotaped Running Biomechanics Analysis. *Phys Med Rehabil Clin N Am, 27*, 217-236.

Souza, R. B., & Powers, C. M. (2009a). Differences in hip kinematics, muscle strength, and muscle activation between subjects with and without patellofemoral pain. *J Orthop Sports Phys Ther, 39*, 12-19.

Souza, R. B., & Powers, C. M. (2009b). Predictors of hip internal rotation during running: an evaluation of hip strength and femoral structure in women with and without patellofemoral pain. *Am J Sports Med, 37*, 579-587.

Taunton, J. E., Ryan, M. B., Clement, D. B., McKenzie, D. C., Lloyd-Smith, D. R., & Zumbo, B. D. (2002). A retrospective case-control analysis of 2002 running injuries. *Br J Sports Med, 36*, 95-101.

Willy, R. W., Manal, K. T., Witvrouw, E. E., & Davis, I. S. (2012). Are mechanics different between male and female runners with patellofemoral pain? *Med Sci Sports Exerc, 44*, 2165-2171.

Willy, R. W., Scholz, J. P., & Davis, I. S. (2012). Mirror gait retraining for the treatment of patellofemoral pain in female runners. *Clin Biomech 27*, 1045-1051.

(1) Conflict of Interest

The authors declare that they have no conflicts of interest in relation to this study.

(2) Ethical Approval

Ethical approval was sought and subsequently granted by the Queen Mary Ethics of Research Committee (QMREC2014/63)

(3) Funding

No funding was received for this project.

Is markerless, smart phone recorded two-dimensional video a clinically useful measure of relevant lower limb kinematics in runners with patellofemoral pain? A validity and reliability study

Abstract

Objectives: Investigate the concurrent validity and reliability of mobile phone collected, high frame rate two-dimensional (2D) video, analysed using the 'Hudl technique' application, compared to three-dimensional (3D) kinematic motion capture during running, in participants with patellofemoral pain (PFP).

Design: Validity and reliability study

Setting: Human biomechanics laboratory

Participants: Males and females with PFP (n=21, 10 males, 11 females, age 32.1 months [±12.9]).

Main Outcome Measures: Manually synchronised 2D and 3D measurement of peak hip adduction (HADD) and peak knee flexion (KFLEX) during running.

Results: 2D and 3D measures of peak KFLEX (p=0.02, *d*=1.13), but not peak HADD (p=0.25, *d*=-0.27), differed significantly. Poor validity was identified for 2D measurement of peak HADD (ICC 0.06, 95% CI -0.35, 0.47) and peak KFLEX (ICC 0.42, 95% CI (-0.10, 0.75). Moderate intra-rater reliability was identified for both variables (ICC 0.61-65), alongside moderate inter-rater reliability for peak KFLEX (ICC 0.71) and poor inter-rater reliability for peak HADD (ICC 0.31).

Conclusions: Measurement of both peak HADD and KFLEX in runners with PFP using mobile phone collected, high frame rate 2D video, analysed using the Hudl Technique Application is invalid, with poor to moderate reliability. Investigation of alternate video analysis approaches to increase precision is warranted. At present, 2D video analysis of running using the Hudl Technique Application cannot be advocated.

Key Words

Patellofemoral Pain, Running, Kinematics, Validity

INTRODUCTION

Recreational running is a common form of exercise (Linton & Valentin, 2018) associated with both positive health benefits (Lee, et al., 2017) and high rates of musculoskeletal injury (19-94%). (Saragiotto, et al., 2014) The knee is reported to be the most prevalent joint involved in running-related musculoskeletal injury, (Linton & Valentin, 2018; Taunton, et al., 2002) with patellofemoral pain (PFP) the most prevalent diagnosis (17%). (Taunton, et al., 2002) New incident cases of PFP amongst recreational runners were recently reported to be 6%. (Bradley S Neal, Lack, et al., 2018)

Whilst musculoskeletal injuries are multi-factorial, (Bittencourt, et al., 2016) peak hip adduction (HADD) during running has been reported as a risk factor for future PFP development in female runners (Noehren, Hamill, & Davis, 2013) and is associated with the persistence of PFP in mixed-sex cohorts. (B. S. Neal, Barton, Gallie, O'Halloran, & Morrissey, 2016) Peak HADD of $\geq$20˚ and a reduction in peak HADD of 5˚ are also reported to be a potential treatment target and mechanism of effect underpinning running retraining in PFP respectively. (Bradley S Neal, Barton, Birn-Jeffrey, Daley, & Morrissey, 2018; Noehren, Scholz, & Davis, 2011; R.W. Willy, Scholz, & Davis, 2012) Peak knee flexion is also a variable of interest in runners with PFP. It is reported to correlate with patellofemoral joint stress (Lenhart, Thelen, Wille, Chumanov, & Heiderscheit, 2014) and is also associated with kinesiophobia, with females with PFP demonstrating lower peak knee flexion angles during stair descent. (de Oliveira Silva, Barton, Pazzinatto, Briani, & de Azevedo, 2016) Altering peak knee flexion may be associated with symptomatic improvements after a step rate retraining intervention. (Bradley S Neal, Barton, et al., 2018)

Guidelines for the measurement of these running kinematics with two-dimensional (2D) video in clinical practice are already in place. (Souza, 2016) Two previous studies have reported concurrent validity and reliability of high frame rate 2D video in comparison to three-dimensional (3D) kinematic motion capture, for measuring peak HADD. (Bart Dingenen, et al., 2017; Maykut, Taylor-Haas, Paterno, DiCesare, & Ford,

2015) Maykut et al. reported a significant, moderate correlation between 2D and 3D measurement for peak HADD during treadmill running (r=0.53-0.62). (Maykut, et al., 2015) In addition, Dinengen et al. reported a significant, positive correlation for peak HADD during over ground running, using a discrete 2D variable to predict an entire 3D kinematic curve from initial ground contact through to toe off. (Bart Dingenen, et al., 2017) Whilst both of these studies reported their methods to be reliable (ICC 0.90-0.99), given that runners with PFP demonstrate differing kinematics compared to matched controls, (B. S. Neal, et al., 2016) the investigation of asymptomatic runners limits the external applicability to clinical populations. Furthermore, both studies analysed their 2D videos using Dartfish software, which may be prohibitive in clinical practice due to high costs and complexity of use relative to simpler, mobile phone-based applications.

Previous studies investigating construct validity for 2D video to measure peak HADD have not identified optimal agreement between 2D and 3D measurement. Ortiz et al hypothesised that transverse plane hip motion may affect the accuracy of 2D measurement of frontal plane hip kinematics during a jump/land task. (Ortiz, et al., 2016) Runners with PFP have also been reported to demonstrate increased peak hip internal rotation (HIR) in comparison to controls. (Noehren, Pohl, Sanchez, Cunningham, & Lattermann, 2012; Noehren, Sanchez, Cunningham, & McKeon, 2012; Souza & Powers, 2009a, 2009b; R. W. Willy, Manal, Witvrouw, & Davis, 2012) Transverse plane motion at the hip is coupled with HADD and tibial abduction, referred to in combination as dynamic knee valgus. (Powers, 2010) Determining the impact of this movement direction on the variability observed between 2D and 3D measurement may provide insight into the source of previously reported sub-optimal agreement. (Ortiz, et al., 2016)

This study aimed to determine whether clinicians can use a simple, readily available tool to measure important lower limb kinematic variables during running. The primary objective was to investigate the concurrent validity and intra- and inter-rater reliability of markerless, high frame rate 2D video, recorded using a smart phone, with reference to 3D kinematic motion capture. The null hypothesis was that smart phone collected 2D video would not give useful measurements of acceptable accuracy with respect to 3D kinematic analyses and as such, a secondary objective was to investigate the source of any identified disagreement.

METHODS

The Queen Mary Ethics of Research Committee (QMREC2014/24/103) gave ethical approval for this study.

Sample size calculation

Using 2D and 3D peak HADD means and a pooled SD from previous work (2D HADD 11.2° [±2.7], 3D HADD 14.0° [3.7]) (Maykut, et al., 2015) and equations for dependent samples t-tests, 21 participants were required to achieve α 5% and β 80% (calculated using G*Power 3.1.9.2, Heinrich-Heine University, Germany). 21 participants with PFP (10 male, 11 female) were conveniently sampled from local sports medicine clinics (see table 1). All participants provided written informed consent prior to participating.

Table 1: Participant characteristics

Variable	**Mean (SD)**
Age (years)	32.1 (12.9)
Height (cm)	169.1 (45.2)
Mass (kg)	69.8 (19.6)
BMI	23.2 (2.6)
Tegner scale	5.5 (1.3)
Symptom duration (months)	53.1 (± 84.5)
Kujala scale	76.2 (± 12.9)
Average NRS	4.7 (± 2.0)

Key: SD=standard deviation; cm=centimeters; kg=kilograms; BMI=body mass index; NRS=numerical rating scale.

Participants

To be eligible, participants were required to have insidious onset retropatellar or peripatellar pain for a minimum of one month, during at least one activity including running, squatting, stair ambulation and jumping. (Crossley, et al., 2016) The Tegner Activity Scale was collected to act as a constant measure across a heterogeneous cohort of participants with PFP who participated in a variety of sports and hobbies.

(Lysholm & Tegner, 2007) Participants below the age of 18 or over the age of 50, or those with traumatic symptoms, patellofemoral instability, tibiofemoral pathology or other concomitant pathology were excluded. Height and mass were collected to allow for the calculation of BMI, reported to be higher in those with persistent PFP. (Hart, Barton, Khan, Riel, & Crossley, 2017) Symptom duration, Kujala scale and average pain over the last 3 months using a numerical rating scale between 0 and 10 (0 = no pain and 10 = worst pain imaginable) were collected as a reflection of symptom severity and persistence, reported to alter running kinematics. (Fox, Ferber, Saunders, Osis, & Bonacci, 2018)

3D kinematics

Kinematic data were collected during running using a four-camera, infrared motion analysis system using Odin software (CX-1, Codamotion, Charnwood Dynamics Limited, Leicestershire, UK), sampling at 200Hz. (Lack, et al., 2014) 24 infrared markers; eight individual markers (10mm) and four rigid clusters of four markers (140mm), were placed adhering to the CAST protocol. (Cappello, Cappozzo, La Palombara, Lucchetti, & Leardini, 1997) Individual markers were placed on the anterior superior iliac spine (ASIS), posterior superior iliac spine (PSIS), lateral calcaneal process and head of 5th metatarsal, with rigid clusters placed on the mid-point of each thigh and shank segment. Foot markers were placed on the participants shoe as an estimation of the anatomical location, given the potential for barefoot running to effect running kinematics. (Hall, Barton, Jones, & Morrissey, 2013) Unpublished intra-rater reliability of kinematic marker placement for the primary investigator (BN) has previously been found to be moderate to excellent (ICC 0.62 – 0.93).

Rigid clusters were secured using a combination of adjustable elastic straps and cohesive self-adherent bandage, with individual markers applied using double-sided adhesive tape and secured with transparent surgical tape. Virtual markers were also identified on the femoral epicondyles and the ankle malleoli, to allow for the calculation of relevant joint centres. The knee joint centre was estimated as the mid-point between the femoral epicondyle markers and the hip joint centre was estimated as a projection within the pelvis frame using previously described methods. (Bell, Pedersen, & Brand, 1990) Joint centre calculation did not differ between male and female participants.

2D kinematics

2D kinematic data were captured using two high frame-rate smartphone cameras (iPhone 6, Apple Corporation, California, USA) recording at 240/frames per second. Cameras were mounted on stable tripods 1.0 metre from the laboratory floor. The camera recording in the sagittal plane was placed at a distance of 2.5 metres from the centre of the ground-embedded force plate (type 9281CA, Kistler Corporation, Switzerland), which participants subsequently ran past. The camera recording in the frontal plane was placed 6.5 metres from the centre of the ground-embedded force plate, which participants ran directly towards (see figure 1).

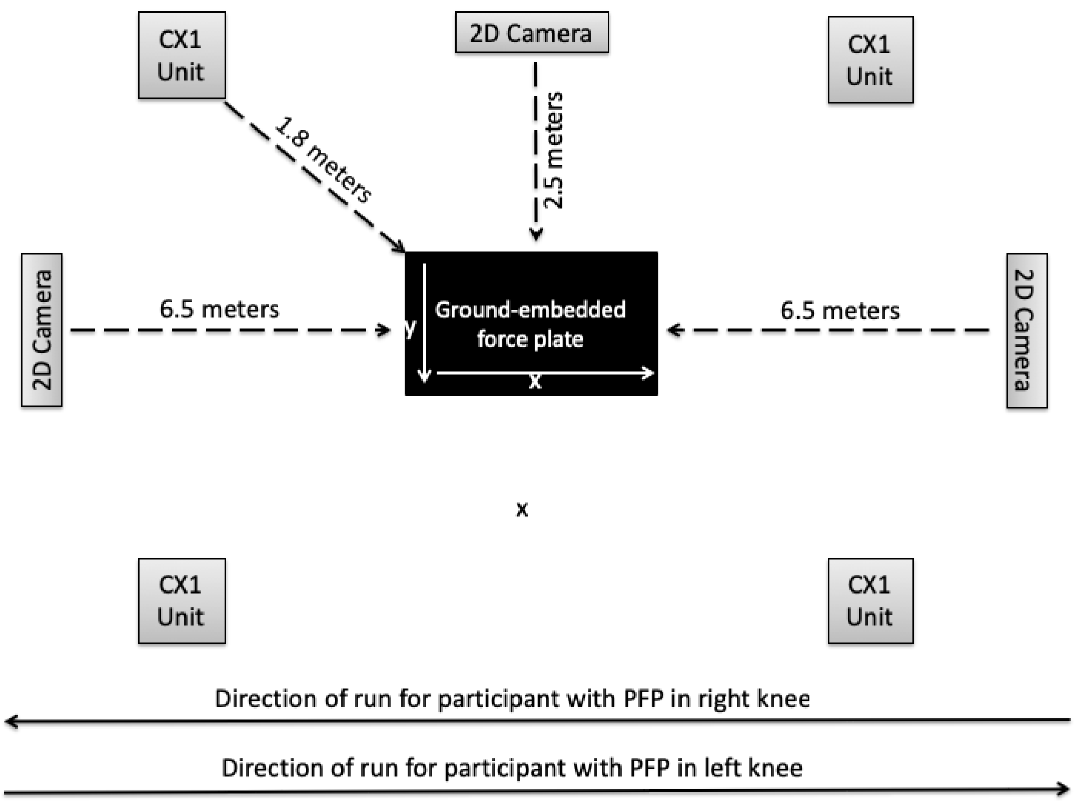

Figure 1: human performance laboratory set up detailing the location of 3D and 2D cameras

Experimental protocol

Both 2D and 3D data were captured during trials of over-ground running in a human performance laboratory. Participants were provided with neutral running shoes in their required size (Asics Nimbus, Asics, Cheshire, UK), to minimise potential effects of footwear variation on running kinematics. (Hall, et al., 2013) Participants were instructed to run in a straight line for a distance of approximately 10.0 meters at a self-selected speed, landing the foot of their symptomatic limb on the ground-embedded force plate, sampling at 1000Hz. The ground-embedded force plate was 5.0 metres from the trial start-point, with participants typically making contact with their fifth step as they ran through. Several practice runs were permitted to allow for familiarisation and to ensure adequate force plate contact during a participant's natural running gait without deceleration. This process was repeated until five successful trials were obtained, with a successful trial defined as an appropriate landing of the correct foot directly onto the force plate without obvious adjustment of running gait. Each trial was initiated by verbal countdown by a member of the research team, with the 3D system and both 2D cameras manually synchronised using a numerical countdown.

Data analysis

To reduce the potential for type I error, data pertaining to one limb only were entered into the analysis. (Menz, 2005) For participants with bilateral symptoms, the limb that rated the highest on the numerical rating scale was evaluated. In the presence of equivalent symptoms the dominant limb was evaluated, defined by the limb that the participant would use to kick a ball.

3D kinematic analysis

Data were analysed offline using a customised Matlab program (version 2015, Mathworks, Natick, Massachussets, USA). A 20*N* threshold from the ground-embedded force plate was used to determine initial contact and toe-off respectively. Kinematic data were processed within this event window, defined as running stance phase. An international society of biomechanics advocated XZY (sagittal, frontal, transverse) cardan rotation sequence was used. Peak joint angles for both peak hip adduction (HADD) and knee flexion (KFLEX) were visualised and subsequently exported to a Microsoft Excel (Microsoft Corporation, Albuquerque, New Mexico, USA) for statistical analysis.

2D kinematic analysis

Videos from successful trials were subsequently imported into the Hudl Technique application (Hudl, Agile Sports Technologies Inc., Nebraska, USA) and analysed independently of the 3D data. 2D data analysis was completed using a tablet device with a 25.9cm screen (5th generation iPad, Apple Corporation, California, USA). Two independent 2D angles, hip adduction (HADD) and knee flexion (KFLEX) were identified using the angle tool. Use of the zoom function within the Hudl technique application was permitted at the discretion of the analyser, to ensure optimal visualization of the relevant anatomical landmarks.

HADD was determined using methods described by Dingenen et al, where the contralateral pelvic drop (CLPD) angle is added from the femoral adduction (FADD) angle. (Bart Dingenen, et al., 2017) CLPD angle was defined as the angle formed by a horizontal line from the stance limb anterior superior iliac spine (ASIS) (referenced from the laboratory floor) and the swing limb ASIS (see figure 2). FADD angle was defined as the angle formed by a horizontal line from the stance limb ASIS (referenced from the laboratory floor) and the centre of the stance limb tibiofemoral joint (an estimation of the knee joint centre) (see figure 2). Within the Hudl technique application, the tool reflects an angle relative to 90˚ and the FADD angle was therefore determined by subtracting the angle produced by the tool from 90˚.

KFLEX was defined as the angle formed by a line drawn from the stance limb greater trochanter to the lateral femoral condyle and a second line drawn from the stance limb lateral femoral condyle to the stance limb lateral malleolus (see figure 2). Within the Hudl technique application, a vertical line in the sagittal plane is reflective of 180˚ and the KFLEX angle was therefore determined by subtracting the angle produced by the tool from 180˚. For both variables, a peak angle was estimated, determined to be when the participant reached the peak of mid-stance, manually defined as the point where maximal foot contact had occurred and no upward/downward motion was occurring. (Maykut, et al., 2015)

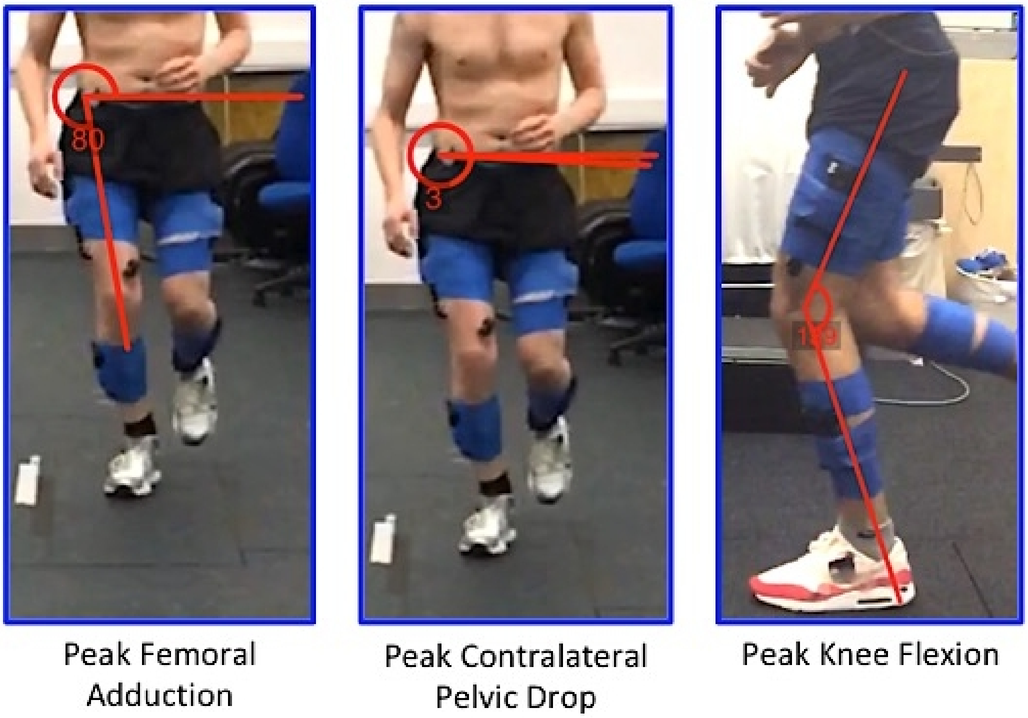

Figure 2: determination of 2D joint angles at the hip and knee

Statistical analysis

All analyses were performed using SPSS (version 22 for MacOS, IBM, New York, USA). The mean of the five 2D and 3D trials were calculated for each participant for both variables of interest (HADD and KFLEX). The difference between the 2D and 3D means was determined using two-tailed, dependent samples t-tests. Single measure ICCs with 95% confidence intervals were calculated using a two-way mixed effects model with absolute agreement, to determine construct validity and both intra- and inter-rater reliability. ICCs were defined as excellent (> 0.90), good (0.75-0.90), moderate (0.50-0.75) and poor (< 0.50) respectively (Koo & Li, 2016). A standard error of measure (SEM) was also calculated to allow for clinical interpretation of reliability data. Bland and Altman plots with 95% limits of agreement (LOA) were used to visually represent the agreement between the 2D and 3D values (Bland & Altman, 1986). Scatter plots were used to visualise the directionality of the relationship between 2D and 3D measurement, with a Pearson's correlation coefficient (*r*) also calculated to allow for comparisons with previous work.

2D peak HADD and KFLEX values from the first run trial of all participants were analysed twice by the primary investigator (BN), with 24 hours between analyses, to determine intra-rater reliability. 2D peak HADD and KFLEX values from the first run trial of all participants were also analysed by a second investigator (SL) and compared to the initial analyses of the primary investigator (BN), to determine inter-rater reliability. Finally, to assess the influence of including 3D peak hip internal rotation (HIR) in a predictive model, a backward linear regression was performed, with the F change statistic used to determine if 3D HIR explains the hypothesised imperfect agreement between 2D and 3D measurement.

RESULTS

Construct validity

There was a significant difference between 2D and 3D measured peak KFLEX, whereas peak HADD was not significantly different between 2D and 3D measures (table 2). ICCs identified a poor correlation for both peak HADD and peak KFLEX between 3D and 2D measurement (table 3).

Table 2: 3D and 2D data for both variables

Variable	3D Measurement (Mean ± SD)	2D Measurement (Mean ± SD)	Difference (Mean ± SD)	*P*	*d*
HADD	12˚ ± 4.7	13˚ ± 3.2	-1˚	0.25	-0.27
KFLEX	38˚ ± 5.5	43˚ ± 3.3	-5˚	<0.01*	-1.13

Key: 3D= three dimensional; 2D=two dimensional; SD=standard deviation HADD=hip adduction; KFLEX=knee flexion.

Table 3: construct validity data for both variables comparing 3D and 2D measurement

Outcome	HADD	KFLEX
ICC (95% CI)	0.06 (-0.35, 0.47)	0.42 (-0.10, 0.75)
Upper LOA	10.9	7.4
Lower LOA	-10.9	-7.4

Key: HADD=hip adduction; KFLEX=knee flexion; SD=standard deviation; ICC=intraclass correlation coefficient; CI=confidence interval; LOA=limits of agreement.

A multiple variable, backward linear regression was calculated to predict 3D peak HADD (dependent variable) using 2D HADD (independent variable$_1$) and 3D HIR (independent variable$_2$). R^2 of the model was 0.06, with a non-significant F change (0.07, p=0.93) identified after the removal of 3D HIR (R^2 change -0.01).

A second multivariable backward linear regression was calculated to predict 3D KFLEX (dependent variable) using both 2D KFLEX (independent variable$_1$) and 3D HIR (independent variable$_2$). R^2 of the model was 0.60, with a non-significant F change (3.76, p=0.06) identified after the removal of 3D HIR (R^2 change -0.08).

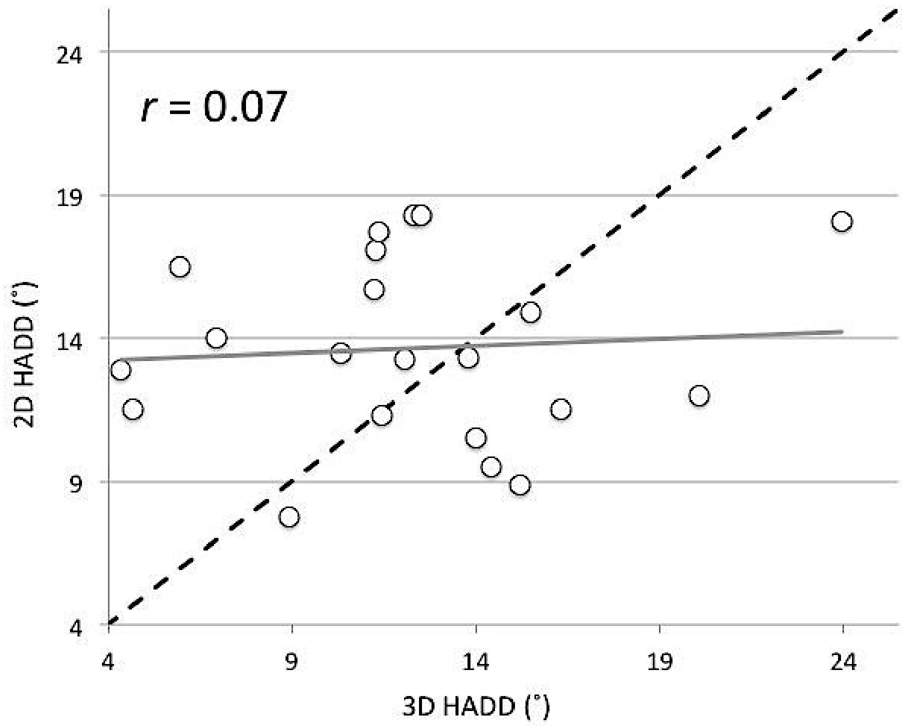

Figure 3: scatter plot for peak 3D and 2D HADD
Key: dashed line represents a line of identity; solid line represents a line of best fit

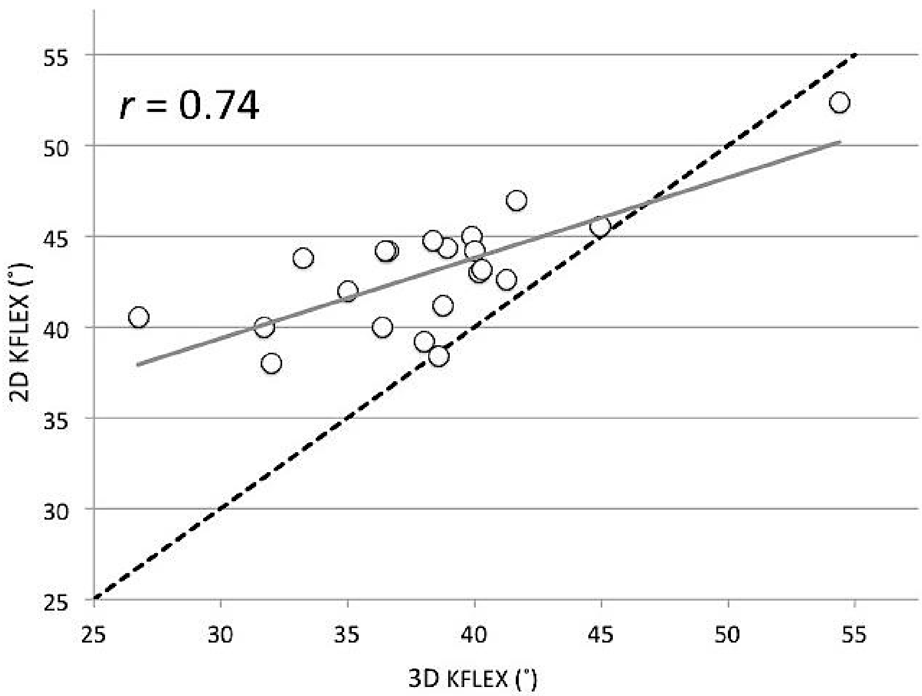

Figure 4: scatter plot for peak 3D and 2D KFLEX
Key: dashed line represents a line of identity; solid line represents a line of best fit

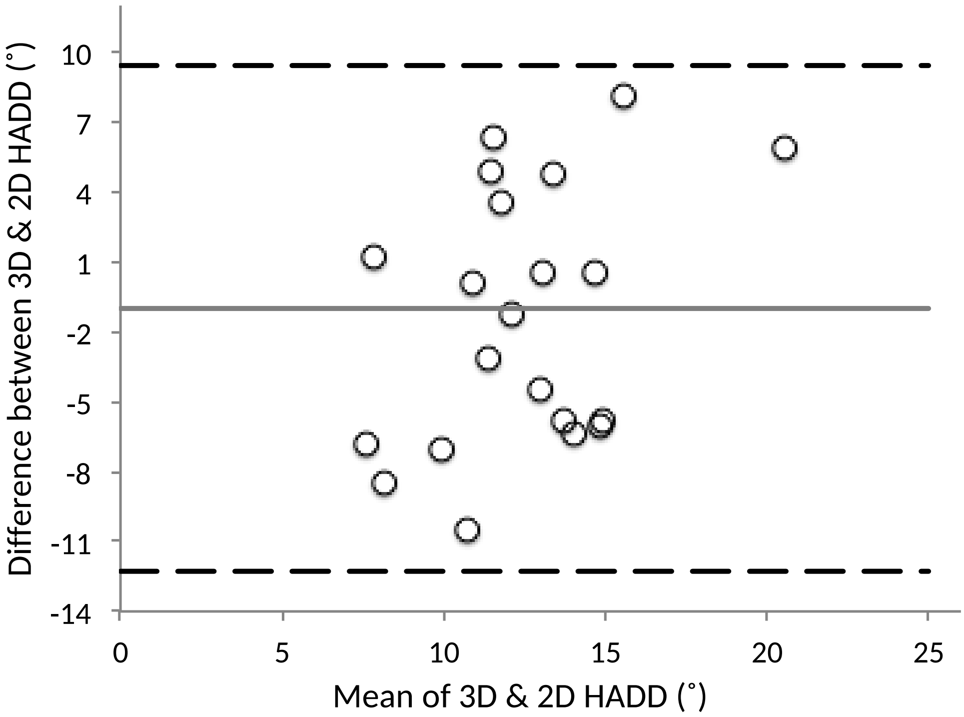

Figure 5: Bland and Altman plot for peak HADD
Key: dashed lines represent upper and lower limits of agreement, solid line represents the pooled mean difference between 3D and 2D measurement.

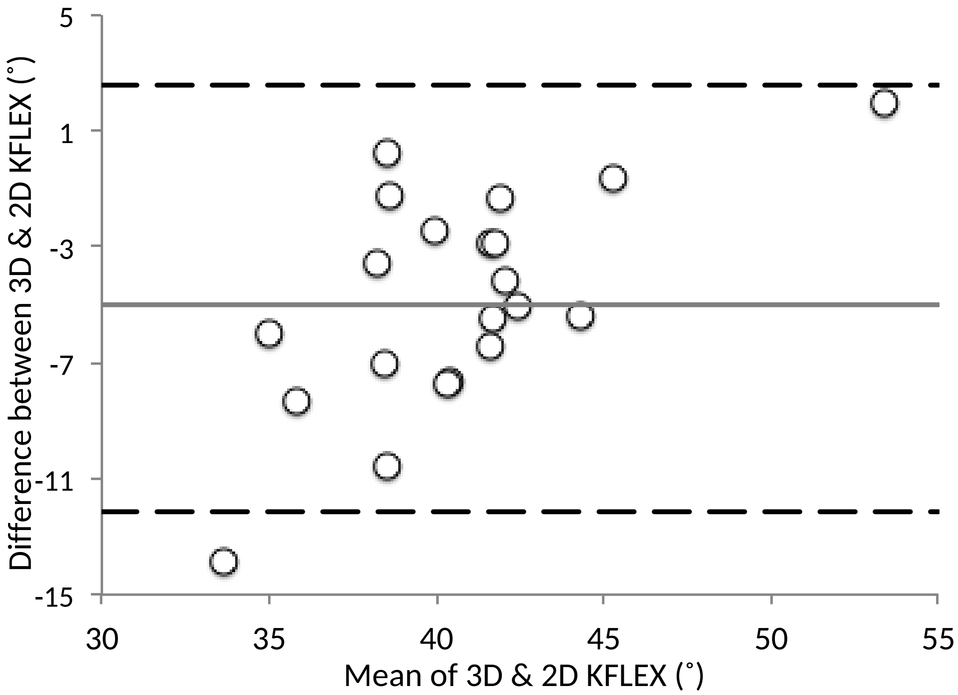

Figure 6: Bland and Altman plot for peak KFLEX
Key: dashed lines represent upper and lower limits of agreement, solid line represents the pooled mean difference between 3D and 2D measurement.

Intra-rater reliability

Moderate intra-rater reliability was identified for peak HADD (ICC 0.65 95% CI 0.34, 0.83, SEM 1.8°) and peak KFLEX (ICC 0.61 95% CI -0.09, 0.87, SEM 2.7°).

Inter-rater reliability

Poor inter-rater reliability was identified for peak HADD (ICC 0.31 95% CI -0.06, 0.64, SEM 3.1°). Moderate inter-rater reliability was identified for peak KFLEX (ICC 0.71 95% CI 0.16, 0.89, SEM 1.4°).

DISCUSSION

Accepting our null hypothesis, 2D measurement of both peak HADD and peak KFLEX was shown to be invalid and have poor to moderate reliability, reflected by low ICCs and wide limits of agreement. These data suggest that markerless, mobile phone collected 2D video, analysed using the Hudl Technique application, does not have acceptable accuracy to quantify either peak HADD or KFLEX during over ground running in individuals with PFP.

Our validity data for peak HADD conflict with the work of both Maykut et al (Maykut, et al., 2015) and Dingenen et al (Bart Dingenen, et al., 2017), who reported significant correlations between 2D and 3D measured peak HADD, despite recording their 2D video at a lower collection frequency. The primary explanation for this disagreement may be the software used to assess the 2D videos. We evaluated the construct validity of the Hudl Technique application, given its ease of clinical application. Hudl Technique is free of cost at the point of access and can be installed on a variety of devices (mobile phones and tablets) and operating systems. The Dartfish software (Dartfish, Fribourg, Switzerland) used in previous studies may offer greater precision, where digitizing 2D video is completed using a mouse on a larger screen, rather than the assessor's finger on a smaller touch screen. The limitation of Dartfish as a method of 2D video analysis is the associated cost (£204-£880 per calendar year).

An additional discrepancy between our study and the work of both Maykut et al (Maykut, et al., 2015) and Dingenen et al (Bart Dingenen, et al., 2017) is the investigation of a cohort of participants with PFP in comparison to asymptomatic participants. Reflective of a typical cohort with persistent PFP (mean symptom duration 53.1 months), our participants had a higher BMI (mean 23.2) than the previously studied asymptomatic cohorts. This may have negatively affected the accuracy of 2D video digitisation by increasing the visual distortion of necessary bony landmarks given the absence of retroreflective markers, particularly the ASIS. Furthermore, our PFP cohort had a lower physical activity level (mean Tegner Scale

5.5) in comparison to the elite asymptomatic cohorts investigated by both Maykut et al (Maykut, et al., 2015) and Dingenen et al (Bart Dingenen, et al., 2017) (estimated Tegner Scale 8-10). Elite runners are reported to have more consistent kinematics than recreational runners (Clermont, Osis, Phinyomark, & Ferber, 2017), which is likely to have resulted in a more stable mean and thus, increased agreement between 2D and 3D measurement (Bart Dingenen, Barton, Janssen, Benoit, & Malliaras, 2018).

A further potential explanation for this conflict is the statistical methodologies employed. Maykut et al (Maykut, et al., 2015) calculated a Pearson's Correlation Coefficient (*r*) which, as a bivariate test, (George, Batterham, & Sullivan, 2003) may over-estimate the agreement between two variables where data demonstrates a linear trend (McGraw & Wong, 1996). This is reflected by the high (*r*) produced by the peak KFLEX data from this study (0.74), versus the low (*r*) produced by the peak HADD data (0.07). Dingenen et al employed statistical parametric mapping (Bart Dingenen, et al., 2017), which does not confirm that the 2D method used can accurately predict a discrete 3D value at a specific point within the gait cycle. Clinicians often seek a discrete kinematic variable within the gait cycle to employ clinical prediction rules, such as a 5° reduction in peak HADD as a predictor for running retraining success, (Noehren, et al., 2011; R.W. Willy, et al., 2012) thus limiting the clinical applicability of these data. A summary of the discrepancies between this study and previous work is presented in table 4.

Table 4: methodological comparison between studies.

	This study	**(Maykut, et al., 2015)**	**(Bart Dingenen, et al., 2017)**
Population	Physically active persons with PFP	Asymptomatic elite runners	Asymptomatic elite athletes
Tegner scale	5	9 (estimated)	9 (estimated)
Mean BMI	23.2	20.0	21.1
Mean age (years)	32.1	19.9	18.7
Running method	Over ground	Treadmill	Over ground
2D video recording frequency	240 fps	60 fps	50 fps
Retroreflective markers	No	Yes	Yes
2D analysis software	Hudl Technique	Dartfish	Dartfish
Statistical method	ICC	Pearson's *r*	SPM
Analysis screen size	Tablet	Computer	Computer
Frontal plane camera distance from axis	6.5 meters	?	4.5 meters

Key: PFP=patellofemoral pain; ?=unable to determine; fps=frames per second; SPM=statistical parametric mapping. Grey shading indicates commonalities between studies.

Our novel investigation of peak KFLEX also demonstrates a poor agreement between 2D video and 3D kinematic motion capture. There is a linear pattern to these data, which results in a Pearson's *r* that over-estimates construct validity (r=0.74 versus ICC=0.42). There also appears to be a systematic bias within these data, with 2D video consistently over-predicting peak KFLEX by a mean of 5˚. Ortiz et al (219) hypothesised that transverse plane hip motion may affect the accuracy of 2D measured running kinematics. Consistent with this hypothesis, there is a statistical trend towards 3D peak HIR being a covariate for this outcome (F change 3.76, p=0.06, R2 change -0.08). Whilst this may explain the systematic bias within these data, this potential model has limited clinical applicability, as transverse plane hip data are not collectable using 2D cameras.

Limitations and future directions

This study is not without limitations, which must be considered when interpreting the results. In an attempt to best replicate clinical practice, participants completed only a short over ground run, with data collected on the fifth step on average. Dingenen et al (Bart Dingenen, et al., 2018) recently reported that a minimum of seven steps are required to allow for a stable mean of a 2D measured kinematic variable. These data refer to analysis completed with Kinovea (http://www.kinovea.org), software that is free of cost at the point of access to Microsoft Windows users. Kinovea offers comparable analysis precision to Dartfish and has been reported to be both inter- and intra-rater reliable for measuring a variety of 2D running kinematic variables (Bart Dingenen, et al., 2018) when data were collected using retroflective markers. Given the apparent potential for increased precision to result in greater construct validity, a future study using either Dartfish or Kinovea involving runners with PFP is warranted.

Only two kinematic variables were assessed in this study and it may be that other kinematic variables prove to be both valid and reliable if investigated by future studies. It could also be that repeating this study using a treadmill running protocol similar to that used by Maykut et al (Maykut, et al., 2015) may return a different outcome. Kinematic comparisons between treadmill and over ground running have been reported to be equivalent (Sinclair, et al., 2013) and a treadmill protocol would allow for the frontal plane camera to be placed closer to the runner, increasing 2D video quality and reducing the potential for parallax error. (B. Dingenen, et al., 2019) Finally, a single video, rather than mean pooled data, were used for the investigation of reliability, differing from the investigation of validity. Whilst this decision was made apriori, analysis of mean pooled data may have yielded different reliability results.

Clinical implications

Whilst the results of this study suggest that markerless, smart phone collected high frame rate 2D video analysed using the Hudl technique application is invalid, there are some implications for clinical practice. Rather than being concerned about maximising video frame rate, attention should be given to placing the 2D camera(s) as close to the runner as possible, to increase quality and reduce parallax error potential. This is most easily achieved using a treadmill rather than over ground running. In addition, use of retroreflective markers is encouraged to maximize ease of identifying relevant bony landmarks, especially those that may be obscured by adipose tissue or clothing. Finally, clinicians are encouraged to analyse 2D data using a large screen and with software that allows for increased precision via use of a computer mouse (or equivalent), rather than a smaller tablet with a touch screen, which is likely to yield inaccurate results.

CONCLUSION

Measurement of both peak HADD and KFLEX in runners with PFP using mobile phone collected, high frame rate 2D video, analysed using the Hudl Technique Application is invalid, with poor to moderate reliability. This may be attributed to the employed 2D video or statistical methodologies, but could also be explained by the increased variability in running kinematics of runners with PFP. Further investigation of methodologies with increased precision is warranted, aiming to improve the ability of high frame rate 2D video to accurately predict 3D kinematics in the clinical setting. At present, clinical gait analysis conducted using the Hudl Technique application should be interpreted with caution, as the validity or reliability of 2D measurement cannot be guaranteed.

Reference List

Bell, A. L., Pedersen, D. R., & Brand, R. A. (1990). A comparison of the accuracy of several hip center location prediction methods. *J Biomech, 23*, 617-621.

Bittencourt, N., Meeuwisse, W., Mendonça, L., Nettel-Aguirre, A., Ocarino, J., & Fonseca, S. (2016). Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *Br J Sports Med*, bjsports-2015-095850.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*, 307-310.

Cappello, A., Cappozzo, A., La Palombara, P. F., Lucchetti, L., & Leardini, A. (1997). Multiple anatomical landmark calibration for optimal bone pose estimation. *Human movement science, 16*, 259-274.

Clermont, C. A., Osis, S. T., Phinyomark, A., & Ferber, R. (2017). Kinematic Gait Patterns in Competitive and Recreational Runners. *J Appl Biomech, 33*, 268-276.

Crossley, K. M., Stefanik, J. J., Selfe, J., Collins, N. J., Davis, I. S., Powers, C. M., McConnell, J., Vicenzino, B., Bazett-Jones, D. M., Esculier, J. F., Morrissey, D., & Callaghan, M. J. (2016). 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 1: Terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome measures. *Br J Sports Med, 50*, 839-843.

de Oliveira Silva, D., Barton, C. J., Pazzinatto, M. F., Briani, R. V., & de Azevedo, F. M. (2016). Proximal mechanics during stair ascent are more discriminate of females with patellofemoral pain than distal mechanics. *Clin Biomech (Bristol, Avon), 35*, 56-61.

Dingenen, B., Barton, C., Janssen, T., Benoit, A., & Malliaras, P. (2018). Test-retest reliability of two-dimensional video analysis during running. *Physical therapy in Sport, 33*, 40-47.

Dingenen, B., Malliaras, P., Janssen, T., Ceyssens, L., Vanelderen, R., & Barton, C. J. (2019). Two-dimensional video analysis can discriminate differences in running kinematics between recreational runners with and without running-related knee injury. *Phys Ther Sport, 38*, 184-191.

Dingenen, B., Staes, F. F., Santermans, L., Steurs, L., Eerdekens, M., Geentjens, J., Peers, K. H., Thysen, M., & Deschamps, K. (2017). Are two-dimensional measured frontal plane angles related to three-dimensional measured kinematic profiles during running? *Physical therapy in Sport*.

Fox, A., Ferber, R., Saunders, N., Osis, S., & Bonacci, J. (2018). Gait Kinematics in Individuals with Acute and Chronic Patellofemoral Pain. *Med Sci Sports Exerc, 50*, 502-509.

George, K., Batterham, A., & Sullivan, I. (2003). Validity in clinical research: a review of basic concepts and definitions☆. *Physical therapy in Sport, 4*, 115-121.

Hall, J. P., Barton, C., Jones, P. R., & Morrissey, D. (2013). The biomechanical differences between barefoot and shod distance running: a systematic review and preliminary meta-analysis. *Sports Med, 43*, 1335-1353.

Hart, H. F., Barton, C. J., Khan, K. M., Riel, H., & Crossley, K. M. (2017). Is body mass index associated with patellofemoral pain and patellofemoral osteoarthritis? A systematic review and meta-regression and analysis. *Br J Sports Med, 51*, 781-790.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med, 15*, 155-163.

Lack, S., Barton, C., Malliaras, P., Twycross-Lewis, R., Woledge, R., & Morrissey, D. (2014). The effect of anti-pronation foot orthoses on hip and knee kinematics and muscle activity during a functional step-up task in healthy individuals: a laboratory study. *Clin Biomech 29*, 177-182.

Lee, D. C., Brellenthin, A. G., Thompson, P. D., Sui, X., Lee, I. M., & Lavie, C. J. (2017). Running as a Key Lifestyle Medicine for Longevity. *Prog Cardiovasc Dis, 60*, 45-55.

Lenhart, R. L., Thelen, D. G., Wille, C. M., Chumanov, E. S., & Heiderscheit, B. C. (2014). Increasing running step rate reduces patellofemoral joint forces. *Med Sci Sports Exerc, 46*, 557-564.

Linton, L., & Valentin, S. (2018). Running with injury: A study of UK novice and recreational runners and factors associated with running related injury. *J Sci Med Sport*.

Lysholm, J., & Tegner, Y. (2007). Knee injury rating scales. *Acta Orthopaedica, 78*, 445-453.

Maykut, J. N., Taylor-Haas, J. A., Paterno, M. V., DiCesare, C. A., & Ford, K. R. (2015). Concurrent validity and reliability of 2d kinematic analysis of frontal plane motion during running. *Int J Sports Phys Ther, 10*, 136-146.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods, 1*, 30.

Menz, H. B. (2005). Analysis of paired data in physical therapy research: time to stop double-dipping? *J Orthop Sports Phys Ther, 35*, 477-478.

Neal, B. S., Barton, C. J., Birn-Jeffrey, A., Daley, M., & Morrissey, D. (2018). The effects & mechanisms of increasing running step rate: A feasibility study in a mixed-sex group of runners with patellofemoral pain. *Physical therapy in Sport*.

Neal, B. S., Barton, C. J., Gallie, R., O'Halloran, P., & Morrissey, D. (2016). Runners with patellofemoral pain have altered biomechanics which targeted interventions can modify: A systematic review and meta-analysis. *Gait Posture, 45*, 69-82.

Neal, B. S., Lack, S. D., Lankhorst, N. E., Raye, A., Morrissey, D., & van Middelkoop, M. (2018). Risk factors for patellofemoral pain: a systematic review and meta-analysis. *Br J Sports Med*, bjsports-2017-098890.

Noehren, B., Hamill, J., & Davis, I. (2013). Prospective evidence for a hip etiology in patellofemoral pain. *Med Sci Sports Exerc, 45*, 1120-1124.

Noehren, B., Pohl, M. B., Sanchez, Z., Cunningham, T., & Lattermann, C. (2012). Proximal and distal kinematics in female runners with patellofemoral pain. *Clin Biomech 27*, 366-371.

Noehren, B., Sanchez, Z., Cunningham, T., & McKeon, P. O. (2012). The effect of pain on hip and knee kinematics during running in females with chronic patellofemoral pain. *Gait Posture, 36*, 596-599.

Noehren, B., Scholz, J., & Davis, I. (2011). The effect of real-time gait retraining on hip kinematics, pain and function in subjects with patellofemoral pain syndrome. *Br J Sports Med, 45*, 691-696.

Ortiz, A., Rosario-Canales, M., Rodriguez, A., Seda, A., Figueroa, C., & Venegas-Rios, H. L. (2016). Reliability and concurrent validity between two-dimensional and three-dimensional evaluations of knee valgus during drop jumps. *Open Access J Sports Med, 7*, 65-73.

Powers, C. M. (2010). The influence of abnormal hip mechanics on knee injury: a biomechanical perspective. *J Orthop Sports Phys Ther, 40*, 42-51.

Saragiotto, B. T., Yamato, T. P., Hespanhol Junior, L. C., Rainbow, M. J., Davis, I. S., & Lopes, A. D. (2014). What are the main risk factors for running-related injuries? *Sports Med, 44*, 1153-1163.

Sinclair, J., Richards, J., Taylor, P. J., Edmundson, C. J., Brooks, D., & Hobbs, S. J. (2013). Three-dimensional kinematic comparison of treadmill and overground running. *Sports Biomech, 12*, 272-282.

Souza, R. B. (2016). An Evidence-Based Videotaped Running Biomechanics Analysis. *Phys Med Rehabil Clin N Am, 27*, 217-236.

Souza, R. B., & Powers, C. M. (2009a). Differences in hip kinematics, muscle strength, and muscle activation between subjects with and without patellofemoral pain. *J Orthop Sports Phys Ther, 39*, 12-19.

Souza, R. B., & Powers, C. M. (2009b). Predictors of hip internal rotation during running: an evaluation of hip strength and femoral structure in women with and without patellofemoral pain. *Am J Sports Med, 37*, 579-587.

Taunton, J. E., Ryan, M. B., Clement, D. B., McKenzie, D. C., Lloyd-Smith, D. R., & Zumbo, B. D. (2002). A retrospective case-control analysis of 2002 running injuries. *Br J Sports Med, 36*, 95-101.

Willy, R. W., Manal, K. T., Witvrouw, E. E., & Davis, I. S. (2012). Are mechanics different between male and female runners with patellofemoral pain? *Med Sci Sports Exerc, 44*, 2165-2171.

Willy, R. W., Scholz, J. P., & Davis, I. S. (2012). Mirror gait retraining for the treatment of patellofemoral pain in female runners. *Clin Biomech 27*, 1045-1051.